

TOOLS AND RESOURCES

A semi-automated machine learning-aided approach to quantitative analysis of centrosomes and microtubule organization

Divya Ganapathi Sankaran^{1,3}, Alexander J. Stemm-Wolf¹, Bailey L. McCurdy¹, Bharath Hariharan^{2,*} and Chad G. Pearson^{1,*}

ABSTRACT

Microtubules (MTs) promote important cellular functions including migration, intracellular trafficking, and chromosome segregation. The centrosome, comprised of two centrioles surrounded by the pericentriolar material (PCM), is the cell's central MT-organizing center. Centrosomes in cancer cells are commonly numerically amplified. However, the question of how the amplification of centrosomes alters MT organization capacity is not well studied. We developed a quantitative image-processing and machine learning-aided approach for the semi-automated analysis of MT organization. We designed a convolutional neural network-based approach for detecting centrosomes, and an automated pipeline for analyzing MT organization around centrosomes, encapsulated in a semi-automatic graphical tool. Using this tool, we find that breast cancer cells with supernumerary centrosomes not only have more PCM protein per centrosome, which gradually increases with increasing centriole numbers, but also exhibit expansion in PCM size. Furthermore, cells with amplified centrosomes have more growing MT ends, higher MT density and altered spatial distribution of MTs around amplified centrosomes. Thus, the semi-automated approach developed here enables rapid and quantitative analyses revealing important facets of centrosomal aberrations.

KEY WORDS: Centrosomes, Microtubules, EB3, Machine learning, Centrosome amplification, Image processing

INTRODUCTION

Microtubules (MTs) are cytoskeletal polymers that perform biological functions essential for life. The interphase MT array is required for cell migration, intracellular trafficking, and cell polarization whereas the mitotic MT array organizes the bipolar spindle and promotes faithful chromosome segregation (Desai and Mitchison, 1997; Hyman and Karsenti, 1996; Inoue and Salmon, 1995). Consisting of polar tubulin subunits, MT polymers have a dynamic plus end and a less dynamic minus end (Allen and Borisy, 1974; Bergen and Borisy, 1980; Desai and Mitchison, 1997; Walker et al., 1988). In cycling cells, minus ends are generally focused at the cell's MT-organizing center or centrosome, where MTs are nucleated and anchored (Brinkley, 1985). Centrosomes consist of a

pair of centrioles surrounded by the pericentriolar material (PCM). Pericentrin and CDK5RAP2 act as scaffolds for recruitment of the γ -tubulin ring complex (γ -TuRC), which nucleates MTs (DICTENBERG et al., 2002; FONG et al., 2007; FARACHE et al., 2018; STEARNS and KIRSCHNER, 1994; MORITZ et al., 1995; ZHENG et al., 1995). During the G1 phase of the cell cycle, cells have one centrosome and two centrioles. Centriole duplication occurs in S phase, resulting in two centrosomes each with two centrioles (HINCHCLIFFE and SLUDER, 2001; PIEL et al., 2000; VOROBYEV and CHENTSOV, 1982). Concurrent to the centriole duplication cycle, PCM organization occurs as a toroid around mature centrioles during interphase, and expands into a more amorphous structure in preparation for mitosis (FU and GLOVER, 2012; LAWO et al., 2012; MENNELLA et al., 2012, 2014; SONNEN et al., 2012; WOODRUFF et al., 2014, 2015). Centrosome duplication doubles the MT nucleation capacity during interphase (SALAYCIK, 2005). Therefore, the number of centrosomes and expansion of the PCM promote the MT architecture of the cell.

Defects in the control of centrosome number commonly occurs in cancer cells, resulting in centrosome amplification (CA) (D'ASSORO et al., 2002; DENU et al., 2016; GANAPATHI SANKARAN et al., 2019; GUO et al., 2007; LOPES et al., 2018; MARTEIL et al., 2018; SALISBURY et al., 2004; SCHNEEWEISS et al., 2003). We and others have previously quantified the changes in centriole number in cancer cells but, to date, how the PCM changes in response to more centrioles remains largely unknown. Moreover, how CA affects MT organization during interphase is not known. This is important because breast cancer therapeutics that target MTs, including Taxol, have been used for decades (PAZDUR et al., 1993; SCHIFF and HORWITZ, 2006), and it highlights the need for quantitative studies of MT organization in normal and breast cancer cells.

Tracking individual MTs within cells is complicated by high MT densities. However, MT organization and dynamics can be measured by quantifying the distribution of fluorescently labeled MTs and/or MT plus-end binding proteins, such as end-binding protein 3 (EB3, also known as MAPRE3), which localizes to growing MT ends (ROTH et al., 2019; SEMENOVA and RODIONOV, 2007; STEPANOVA et al., 2003; STRAUBE and MERDES, 2007; APPLIGATE et al., 2011). Counting EB3 foci, or comets, is time-consuming and requires quantitative computational approaches (Applegate et al., 2011). Moreover, these tools require extensive manual annotation and the detection of centrosomes. Annotating each centrosome is a laborious process that often requires confirmation of colocalization between signals by toggling between image channels and manually adjusting brightness and contrast. Automation of this analysis requires the automated detection of centrosomes, the identification of EB3 comets around detected centrosomes and the analysis of EB3 comet distributions. Similar approaches are also required to quantify the distribution of MTs surrounding centrosomes.

¹Department of Cell and Developmental Biology, University of Colorado School of Medicine, 2801 East 17th Avenue, Aurora, CO 80045-2537, USA. ²Department of Computer Science, Cornell University, Ithaca, NY 14853, USA. ³Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, USA.

*Authors for correspondence (bharath.hariharan@cornell.edu; Chad.Pearson@cuan.schutz.edu)

 C.G.P., 0000-0003-1915-6593

Handling Editor: Guangshuo Ou
Received 3 January 2020; Accepted 18 June 2020

In this study, we establish a semi-automatic analysis pipeline that uses machine learning to automatically detect centrosomes followed by user-defined correction of errors. Machine learning is a class of techniques where a computational model is ‘trained’ to emulate the annotations of an expert. The problem of detecting centrosomes introduces additional challenges to extant machine learning pipelines: the number of training annotations is too small to train existing machine learning models, and the large size of the image files makes existing models run slowly. Here, we introduce a new pipeline that tackles these challenges so that training is efficient and the run time is quick. The identification of centrosomes by machine learning is then coupled to automated analysis pipelines for the PCM, MTs and EB3 comets around the identified centrosomes. Finally, the entire pipeline is encapsulated in a new graphical tool that allows users to visualize the automated detections and to correct errors from the automated analysis.

This semi-automated image analysis pipeline was used to investigate centriole and centrosome frequency, PCM size and MT distributions in normal and breast cancer cells with amplified centrosomes. We discovered that breast cancer cells with amplified centrosomes exhibit increased PCM protein levels and PCM size, based on γ -tubulin and pericentrin fluorescence. Moreover, we found that breast cancer cells with amplified centrosomes have elevated MT density and MT growth near centrosomes. In summary, centrosome amplification increases MT nucleation and promotes changes to MT density and spatial distribution of MTs around amplified centrosomes in breast cancer cells.

RESULTS

Machine learning for centrosome detection and cell segmentation

As a first step in creating an automated analysis pipeline for determining changes to the MT organization in centrosome-amplified breast cancer cells, we developed a centrosome detector using machine learning. Centrosome foci can be difficult to detect because of large variations in the signal-to-noise ratio. In preliminary experiments, we found that simple image processing techniques that rely on consistent signal-to-noise ratios often failed to detect these foci accurately. An alternative algorithm for detecting centrosomes was designed by adapting existing object detection protocols using computer vision (Liu et al., 2016). This prior work uses convolutional networks to assign a score to every location in the image. The core output at a particular location is interpreted as the probability of a centrosome at that location. Convolutional networks are machine learning models comprised of sequences of convolution operations interspersed with subsampling operations. The filters of these convolutions are automatically estimated by the learning algorithm based on a training dataset consisting of pairs of images and the desired outputs. We ‘train’ the convolutional network on a dataset consisting of images with annotated locations of centrosomes. For each centrosome, we annotate the two centrioles, which are marked only when the centrin foci colocalize with the PCM. Thus, the convolutional network is trained to detect centrosomes by localizing the two centrioles.

We found that using prior out-of-the-box convolutional network architectures used in computer vision presented two major challenges. First, these models require millions of training images that are not pertinent to biological samples (He and Sun, 2014). This is because the convolutional network has a large number of parameters; with more parameters, more data are required to optimize these parameters. Second, these architectures are designed to run on small images (typically 224×224 pixels) and so are very

slow when run on large fluorescence microscopy images (typically greater than 1000×1000 pixels).

To address these challenges, we designed a new convolutional network architecture that requires less training data, memory and time to run (Fig. 1A). The input to the network is a three-dimensional fluorescence image stack where each xy coordinate is projected to their maximum pixel intensity. We utilized maximum intensity projections because analyzing complete three-dimensional volumes is slow and difficult to train owing to the large number of pixels that need to be analysed, with a vanishingly small fraction of pixels that will correspond to centrosomes. The mean and standard deviation of the fluorescence intensity across all pixels is calculated, and the mean is subtracted from each pixel and divided by the standard deviation. Subtracting the mean makes the result invariant to changes in brightness in the original image or, more precisely, the addition of a constant to the intensity of all pixels. Analogously, dividing by the standard deviation makes the result invariant to changes in contrast or, more precisely, multiplication of the intensity of all pixels by a constant. The result is thus brightness- and contrast-normalized.

The network then uses this normalized intensity image to compute four images (corresponding to four scores per pixel; Fig. 1A, panels 2–5). Two of these images are the fluorescence intensity for centriole (Fig. 1A, panel 3) and PCM markers (Fig. 1A, panel 4); we used these because colocalization of high-intensity foci in these channels likely indicates centrosomes. By themselves, these channels might be noisy and produce spurious foci. We obtained two more images by running a small, fast four-layer fully convolutional network on each individual channel (Fig. 1A, panels 2 and 5). The network reduces the resolution by a factor of eight, but also reduces noise and removes spurious foci. Reducing resolution has two benefits. First, it makes the network much faster. Second, because the convolution operations used by the network operate on fixed-size image patches, at a lower resolution these patches correspond to a larger fraction of the image, thus allowing the network to analyze a larger part of the image to make decisions for each pixel. All outputs are up-sampled using nearest-neighbor interpolation to the size of the original image and converted to a score between 0 and 1 using the sigmoid function:

$$f(x) = \frac{1}{1 + e^{-x}}$$

before being multiplied together, resulting in a final accurate score for each pixel corresponding to the likelihood of finding a centrosome at that location (Fig. 1A, panel 6). Our model applies this sigmoid function on four different channels (x) (Fig. 1A). Two of those channels are outputs from a small convolutional network, which can learn to scale its outputs appropriately during training. The other two channels are the centrin and pericentrin, which are not scaled beyond the brightness and contrast normalization mentioned above. The full model has fewer than 10,000 parameters and is very efficient (fewer than 7 gigaflops, compared to 35 gigaflops for a standard object detection approach; Liu et al., 2016).

Local maxima in this output were identified as centrioles colocalized with PCM (circles in Fig. 1A, panel 7). More precisely, pixels in the output were considered sequentially in decreasing order of scores. The highest scoring pixel was declared as a centriole, and subsequent, lower-scoring pixels were declared centrioles only if no previously declared centriole was fewer than $r=5$ pixels (0.65 μm) away (this is called non-maximum suppression; Viola and Jones, 2011). This produces a ranked list of putative centrosome detections, and the user can choose where to threshold this list.

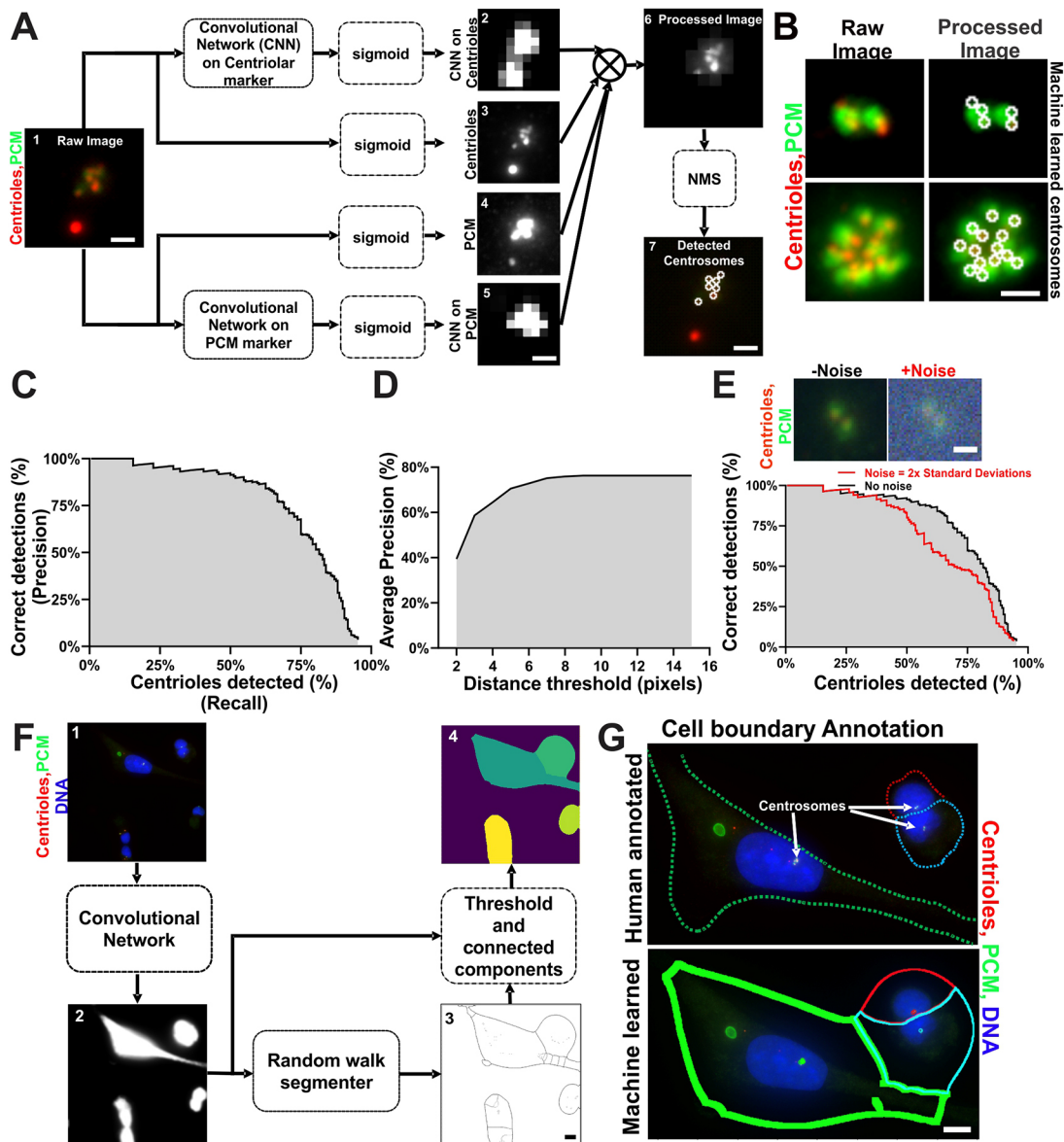


Fig. 1. A machine learning algorithm for centrosome and cell detection. (A) Separate convolutional networks operating on centrin (centrioles, red) and pericentrin (PCM, green) fluorescence signals assign a score for each pixel indicating the likelihood that it is a centrosome (panels 2 and 5). These are passed through a sigmoid function $[f(x)=1/(1+e^{-x})]$ and multiplied together with the intensities of the centrin and pericentrin channels (panels 3 and 4), also after a sigmoid function, to produce a final score for each pixel (panel 6). Peaks in the final score are detected using non-maximum suppression (NMS) and identified as centrosomes (panel 7). Scale bars: 1 μ m. (B) Left panels, raw images of centrosomes stained for centrioles (centrin, red) and PCM (γ -tubulin, green). Right panels, detected centrioles (white circles). Scale bar: 1 μ m. (C) Evaluation of the centriole detection in terms of the fraction of centriole detections that are deemed correct (precision) and the fraction of centrioles detected (recall). A detection is deemed correct if it falls within a distance of ten pixels from the correct centriole. The plot shows precision and recall as the selection criterion is made less conservative by reducing the threshold score at which a centriole is detected. (D) Variation of average precision (averaged over multiple recall values from 0 to 100% as the criterion for a correct detection is made more lenient. The x-axis is the maximum distance between a true and a predicted centriole at which the predicted centriole is still considered correct). (E) Precision and recall values for a noisy image compared to the original image. Staining: centrioles (centrin, red) and PCM (γ -tubulin, green). Scale bar: 1 μ m. (F) Pipeline for segmenting out individual cells. (1) original image, (2) output of the neural network that identifies pixels that fall within cells, (3) boundary map obtained from the random walk segmenter with estimated boundary strength, (4) final segmentation. Staining: centrioles (centrin, red) and PCM (γ -tubulin, green). Scale bar: 10 μ m. (G) Comparison between predicted cell segmentation and human-annotated cells. Staining: centrioles (centrin, red) and PCM (γ -tubulin, green). Scale bar: 10 μ m.

The non-maximum suppression algorithm described here is commonly used in the computer vision literature and has been proven to be more effective than most other alternatives (Girshick et al., 2014). While in principle the algorithm can potentially identify pixels that are not local maxima in large regions of almost uniform score, in practice this is unlikely to happen because the convolutional network is strongly penalized during training if it produces such regions. The full centrosome

detection pipeline is shown in Fig. 1A and example detections are shown in Fig. 1B.

We annotated centrosomes on a small dataset of ten images and used these as training images to train the convolutional networks. To evaluate this centrosome detection approach and make sure that it was indeed detecting centrosomes correctly, we conducted the following ‘leave-one-out’ evaluation. We obtained algorithmic detections for each annotated image using a model trained on the

other annotated images. We then marked detected foci as correct if they were within 10 pixels (1.3 μm) of a human-annotated centriole, and as spurious otherwise. If multiple foci were detected where there was only one centriole, all but one of the foci were considered spurious. We then measured the precision (the fraction of detected foci that were deemed correct) and the recall (the fraction of human-annotated centrioles that were detected). An ideal algorithm would detect all and only correct centrosomes, achieving precision and recall of 100%. We plotted how precision varied with recall when the score threshold for declaring a centrosome was reduced (Fig. 1C). Although not perfect, the centrosome detector maintained a precision of 90% when recall was close to 50%, and still maintained a precision of more than 50% at high levels of recall (>75%). We compared the model's results with that of an alternate baseline model that did not have any convolutional network on either channel. This baseline model only used channels 3 and 4, as shown in Fig. 1A, and introduced significantly more false positives. As a result, even when detecting less than 30% of the centrosomes, more than 10% of its detections were false positives (Fig. S1B).

We next evaluated the ability of the centrosome detector to accurately localize the centrosome (Fig. 1D). As above, detected foci were marked correct if they were within a threshold distance of a human-annotated centriole. We varied this distance threshold, and at each distance threshold we computed the average precision, or the precision averaged over multiple values of recall. We found that the centrosome identification maintains a high average precision even for stringent thresholds (a distance of 5 pixels corresponds to 0.325 μm and 0.65 μm , through 2 \times 2 binning), indicating that the detector accurately localizes the centrioles of centrosomes. Finally, we asked how resilient the centrosome detection is to noise. We artificially added Gaussian noise to the image and computed the precision versus recall curve for the centrosomes identified in a noisy image (Fig. 1E). Even with noise more than 16 times the standard deviation of the original image, centrosomes were consistently detected. However, note that the assumption of Gaussian noise might not correspond to the noise observed in real microscopy images.

Importantly, although the centrosome detector does make errors, the accuracy is a function of the training data. To understand how additional training data might improve accuracy, we annotated an additional ten images, and compared the accuracy of a model trained on five images with a model trained on 15 images (Fig. S1E). We found that the additional training data significantly improved accuracy, both in terms of precision (i.e. it reduced the incorrect detections) and the recall (i.e. the model trained on the larger training set detected more centrosomes).

Next, the analysis requires that individual cells are separately analyzed. To group the detected centrosomes into each individual cell, we created a pipeline for segmenting the cells in the image (Fig. 1F). In the first step, a convolutional network identifies all pixels that fall inside a cell. This convolutional network is trained using a small dataset where the cells have been annotated by hand. The output of this network can be interpreted as a probability $p(x)$ for each pixel x that indicates whether it falls inside a cell. The next step is to group the pixels with a high $p(x)$ into separate cells. To do this, local peaks were identified in this output as markers for possible cells and a random walk segmenter (Grady, 2006) was used to segment the cell pixels by assigning each cell pixel to one of the markers. However, the random walk segmenter can over-segment and subdivide a single cell into multiple cells. We, therefore, estimate the strength of the boundary between the cells using the convolutional network output. We define this boundary strength as:

$$\sum_{i=1}^n (1 - p(x_i)) / n,$$

where the summation is over all pixels on the cell boundary (Fig. S1A). A user-defined threshold was used to merge cells that are separated by weak boundaries. We qualitatively compared the result of the semi-automatic segmentation approach to a human-annotated cell. Although the machine-generated segmentation does not accurately track cell boundaries, it was able to identify all the cells and capture the bulk of the interior of the cell, especially in the vicinity of the centrosome (Fig. 1G).

Taken together, these results suggest that our centrosome detection and cell segmentation approach can increase the speed and accuracy of analyses but may require manual intervention when predictions are incorrect.

Evaluation of PCM and MT organization using machine learning-aided image processing

Once centrosomes and cell boundaries have been detected, we designed automated image processing algorithms to analyze the spatial distribution of centrosomal PCM proteins, MTs and EB3 comets at and around the centroid of centrosomes (Fig. 2A–D). This analysis routine can be automatically run on cells with detected centrosomes. These image processing routines and the machine learning-aided detection of centrosomes and cells described above were encapsulated into a semi-automatic graphical tool for analysis.

Below we first describe the image processing algorithms for the analyses of centrosomal proteins, MTs, and EB3 comets. We then describe the semi-automatic graphical tool that encompasses both these image processing algorithms and the machine learning-aided detection of centrosomes and cells.

PCM and MT intensity analyses

To quantify PCM and MT intensity in concentric rings at and around centrosomes, images were projected to their maxima and the centroid of the identified centrosomes was algorithmically computed. When there are multiple centrosomes that are spatially separated by more than 40 pixels ($\sim 5 \mu\text{m}$), the algorithm groups them into separate clusters using single-link hierarchical clustering and performs the analysis only on the largest cluster. Next, the distance of each pixel from the centrosome centroid in two dimensions was calculated (Fig. 2A, top panel). For each radius from inner r_1 to r_n where n is the user-defined number of radii used in the analysis, the tool computes the total fluorescence intensity (I_n) of the pixels between the concentric radii. The difference in total intensity ($I_{n+1} - I_n$) between the n th and $(n+1)$ th regions provides the fluorescence intensity in the n th concentric ring [Fig. 2A, middle panel; Fig. 2B, top panel; n th, orange ring; $(n+1)$ th, blue concentric ring]. To correct for the total area of each ring the intensity is normalized to the number of pixels within each ring.

Only pixels that fall within the cell boundary are quantified. To define the cell boundary for the MT intensity computation, a low threshold was used on the α -tubulin fluorescence channel; this was motivated by the fact that microtubule fluorescence extends throughout the cell area (Fig. 2B, bottom panel). In a case where such a channel is not available, one can also use the cell body estimated by the cell segmentation algorithm as a rough estimate. For the PCM density computation, smaller radii (in steps of 0.13 μm) and a smaller total diameter were used so that the analysis region was always inside the cell. Cells with centrosomes close to the edge of the cell or the edge of the image were not analyzed. Therefore, we did not need to account for the cell area (Fig. 2A). In each case, once the pixels within the cell were defined (i.e. the

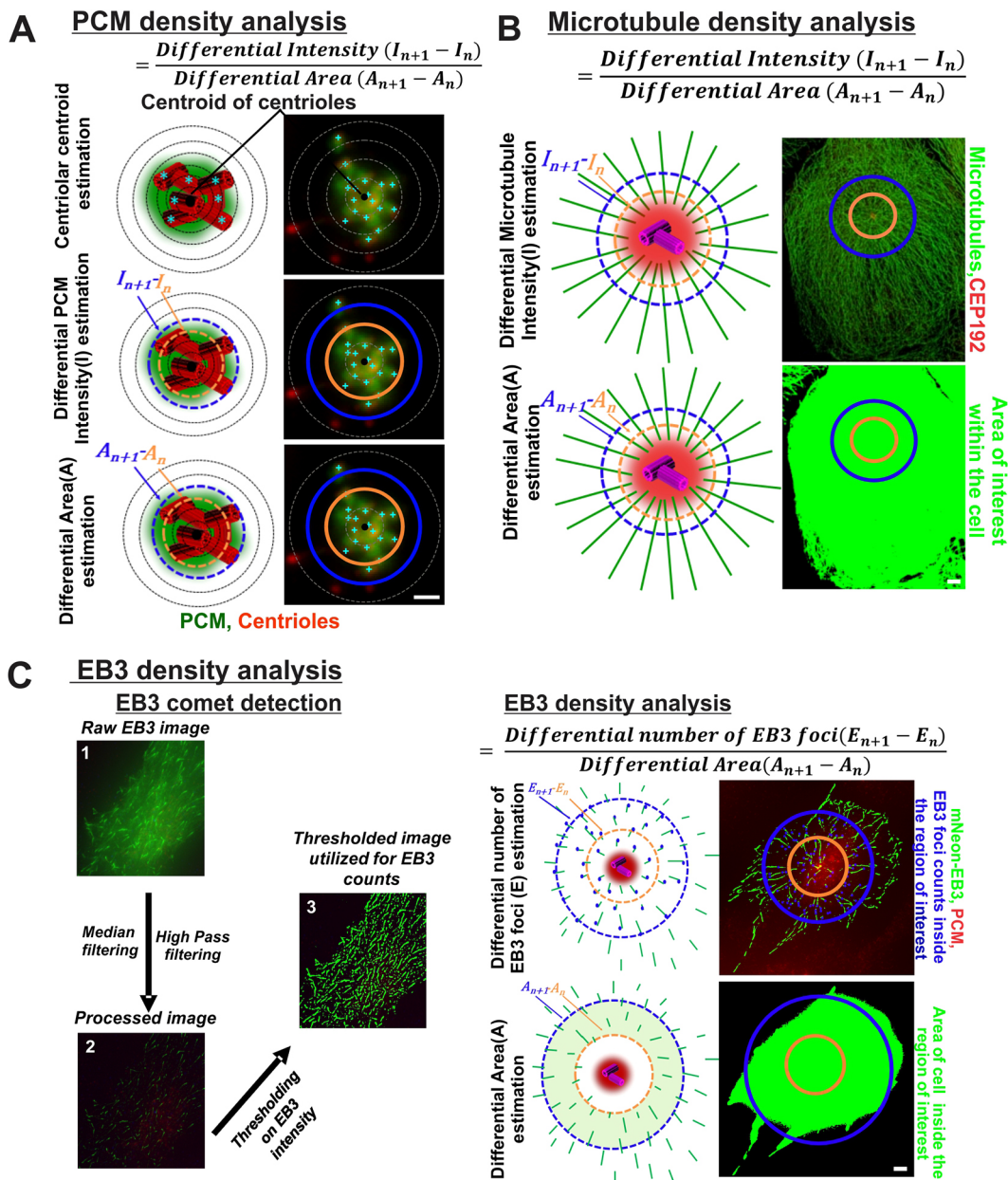


Fig. 2. Algorithms to quantify PCM and microtubule density around centrioles. (A) PCM density analysis. Top left panel, schematic of centriolar centroid (black circle) estimation based on centrioles (cyan foci). Top right panel, centrosomes in MDA-231 cells stained for centrioles (centrin, red) and γ -tubulin (PCM, green). Middle left panel, schematic for analysis of PCM intensity. Middle right panel, the region around the centroid of the centrioles is divided into concentric rings, and the differential intensity ($I_{n+1}-I_n$) of pericentriolar material is calculated around centrosomes. Bottom left panel, schematic for analysis of differential area. Bottom right panel, the region around the centroid of the centrioles is divided into concentric rings, and the differential area ($A_{n+1}-A_n$) is calculated around centrosomes. Scale bar: 1 μ m. (B) MT density analysis. Top left panel, schematic of analysis of differential MT intensity ($I_{n+1}-I_n$) for each ring. Top right panel, centrosomes in MDA-231 cells stained for centrioles (CEP192, red) and MTs (α -tubulin, green). Bottom left panel, schematic for analysis of differential area ($A_{n+1}-A_n$) around centrosomes in MDA-231 cells. Bottom right panel, image of thresholded MTs. Scale bar: 5 μ m. (C) EB3 density analysis. Left, the pipeline for detecting EB3 foci. (1) Original image (mNeon-EB3, green). (2) Processed image after applying the median filter and subtracting a box filter blurred version of the same image. Pixels on EB3 foci have high intensity and all other pixels are made close to 0 intensity. (3) Thresholded binary image. Connected components in this image are considered as EB3 foci. Right, schematic of analysis of EB3 density for each ring. Top left panel, schematic for analysis of differential EB3 counts ($E_{n+1}-E_n$). Top right panel, centrosomes in mNeon-EB3 MCF10A cells stained for centrioles (centrin, green) and γ -tubulin (red). Bottom left panel, schematic for analysis of the differential area. Bottom right panel, image of thresholded EB3 foci. Scale bar: 5 μ m.

cellular area) then, for each radius from inner r_1 to r_n the tool computes the differential area ($A_{n+1} - A_n$) between the n th and $(n+1)$ th regions. Dividing the differential intensity by the differential cell area provided the density (Fig. 2B). Altogether, this generated an analysis routine to measure the spatial distribution of MTs and PCM fluorescence per unit area.

EB3 Foci analysis

To quantify the distribution of MT nucleation and growth from centrosomes, we designed a similar algorithm to measure the distribution of EB3 foci on images of fixed cells (Fig. 2C). The algorithm detects EB3 comets by utilizing a median filtering step to remove noise in the form of stray pixels with high fluorescence

intensity. The median filter replaces each pixel with the median of its neighbors, thus removing pixels having very high or very low intensity values. Then, to identify pixels with greater intensity than their neighbors, high-pass filtering was performed. This involves first computing a blurred version of the image by convolving with a box filter, which replaces every pixel by the average of its neighbors in a $k \times k$ neighborhood, and then subtracting the blurred image from the original. Thresholding this differential image provided a binary image that identified pixels corresponding to EB3 comets. Finally, connected component analysis was performed on this binary image. This analysis identifies contiguous regions of the image that have a 'true' value. Each connected component identified in this binary image was considered an EB3 comet (Fig. 2C).

Analyzing the distribution of EB3 comets in concentric rings, as described above, requires additional considerations because EB3 comets can straddle multiple rings. For each EB3 comet, we identified the pixel on the comet farthest from the centroid of the centrosomes. Then, for each radius (r_n) from the centroid of the centrosome, we counted the EB3 comets for which the most centrosome-distal pixel fell within this radius. The differential EB3 counts were then computed in the n th ring as the difference in comet counts ($E_{n+1} - E_n$) and divided by the differential cell area to obtain the EB3 density (Fig. 2C, right panel).

A semi-automatic graphical tool for MT, PCM and EB3 analyses

We encapsulated the centrosome detection, cell segmentation, and MT, PCM and EB3 analysis algorithms described above into a single tool with a graphical user interface (Fig. 3A). This graphical user interface does not include the 'training' of the machine learning-based centrosome detector, which is trained separately and used for all subsequent analyses.

To validate the performance of the machine learning-based centrosome detector, we compared its detections to human annotation. For this comparison, we used the output of the machine learning-based centrosome detector without any corrections, except for a tuning of the threshold on the score for each image. We first compared the number of interphase cells with centrosomal annotations identified by the machine versus the number identified by human annotation. The average number of interphase cells with centrosomal annotations detected in a replicate by the machine was 52, which is very similar to the average of 55 identified by humans (Fig. 3B). Furthermore, the numbers of human- versus machine-annotated, centrosome-amplified and non-amplified cells were similar (Fig. 3C). We next assessed the number of centrioles detected by the machine and compared it to the number detected by human annotation. Most annotations fall on the identity line (the line of unit slope) (Fig. 3D). This suggests that most machine-annotated and human-annotated centriole numbers were similar (Fig. S1C). Linear regression on centriole numbers also suggested a line with approximately unit slope (1.12). Although some annotations fall outside the identity line, they are present on both sides of the identity line suggesting that there was not a bias for too few or too many centrioles in the prediction by the algorithm. Furthermore, on mitotic cells, the algorithm detected both centrosomes accurately. The percentage of mitotic centrosomes detected by the machine was similar to that identified by humans (Fig. 3E,F; Fig. S1D). These data validate the performance of the machine-learning centrosome detector.

To allow for corrections to the machine output, this semi-automatic tool permits manual intervention. In particular, the user can choose thresholds for centrosome detection and cell segmentation and visualize the machine's outputs. The user can

correct these outputs by removing spurious foci, adding centrosomes and cell boundaries that were not detected and merging cells that have been mistakenly marked as two separate cells. The tool also allows the user to control various aspects of the analysis itself, such as the radii used for spatial analysis, the thresholds for identifying EB3 comets, and the thresholds for determining the cell area. Finally, this tool can be easily generalized to analyze the distribution of any protein around centrosomes or any other user-identified foci.

With this tool, analysis is performed with the following steps. First, the centrosome detection model and cell segmentation model are loaded followed by loading of the image (Fig. 3A, step 1). Second, the centrosome detection and cell segmentation models run [run machine learning (ML) models; Fig. 3A, step 2]. This automatically detects centrosomes and cell boundaries and presents them as an overlay on the original image. In case of an error in the cell segmentation, the segmentation can be corrected by adding additional cell boundaries or merging wrongly segmented cells. One can also remove cells from further analysis (Fig. 3G, step 3). The tool also allows for the removal of spurious detections or the inclusion of missed centrosomes (Fig. 3H, step 4). The tool automatically characterizes cells as being centrosome-amplified based on the number of detected centrosomes (cells with greater than two centrosomes), with amplified cells highlighted by a thick outline. This automated labeling can also be corrected if needed (Fig. 3A, step 5). Finally, once the centrosomes have been finalized, the type of analysis is chosen (Fig. 3A, step 6). The analysis is performed automatically, and the results saved in a tabular format.

The tool also allows for corrections (i.e. removal of spurious detections and identification of missed centrosomes) to be saved for future training of the centrosome detection model. The training can be performed using a separate command-line utility that is packaged with the tool.

In summary, we have generated a new tool with a graphical interface that allows users to analyze the spatial distribution of centrosomal proteins and MTs around centrosomes. The tool performs semi-automated analyses, automating parts of the analysis pipeline but also allowing for user interventions.

The semi-automatic machine learning algorithm detects PCM defects in breast cancer cells

To test whether PCM levels are elevated in breast cancer cells with amplified centrosomes, the algorithm described above was used to quantify the levels and distribution of PCM proteins in normal mammary epithelial cells and breast cancer cells (MCF10A and MDA-231, respectively).

Analyses were first performed using manual annotation of centrosomes. The levels and distribution of PCM proteins (γ -tubulin and pericentrin) were quantified at centrosomes with normal or amplified numbers. Approximately 23% of MDA-231 breast cancer cells had CA, whereas only 5% of normal-like breast cells (MCF10A) had CA (Ganapathi Sankaran et al., 2019). The relative intensities of γ -tubulin and pericentrin were quantified per unit area in normal and amplified centrosomes (Fig. 4). Centrosomal γ -tubulin was elevated by $\sim 40\%$ at amplified centrosomes compared to non-amplified centrosomes in both MCF10A and MDA-231 cells (Fig. 4A,B). We next asked whether the relative γ -tubulin intensity per centrosome is different between a non-amplified and an amplified centrosome. In order to answer this, we arrested MDA-231 cells in S phase and quantified the relative intensity of γ -tubulin per centrosome per unit area. We found that the centrosomal γ -tubulin was elevated by $\sim 30\%$ in an amplified

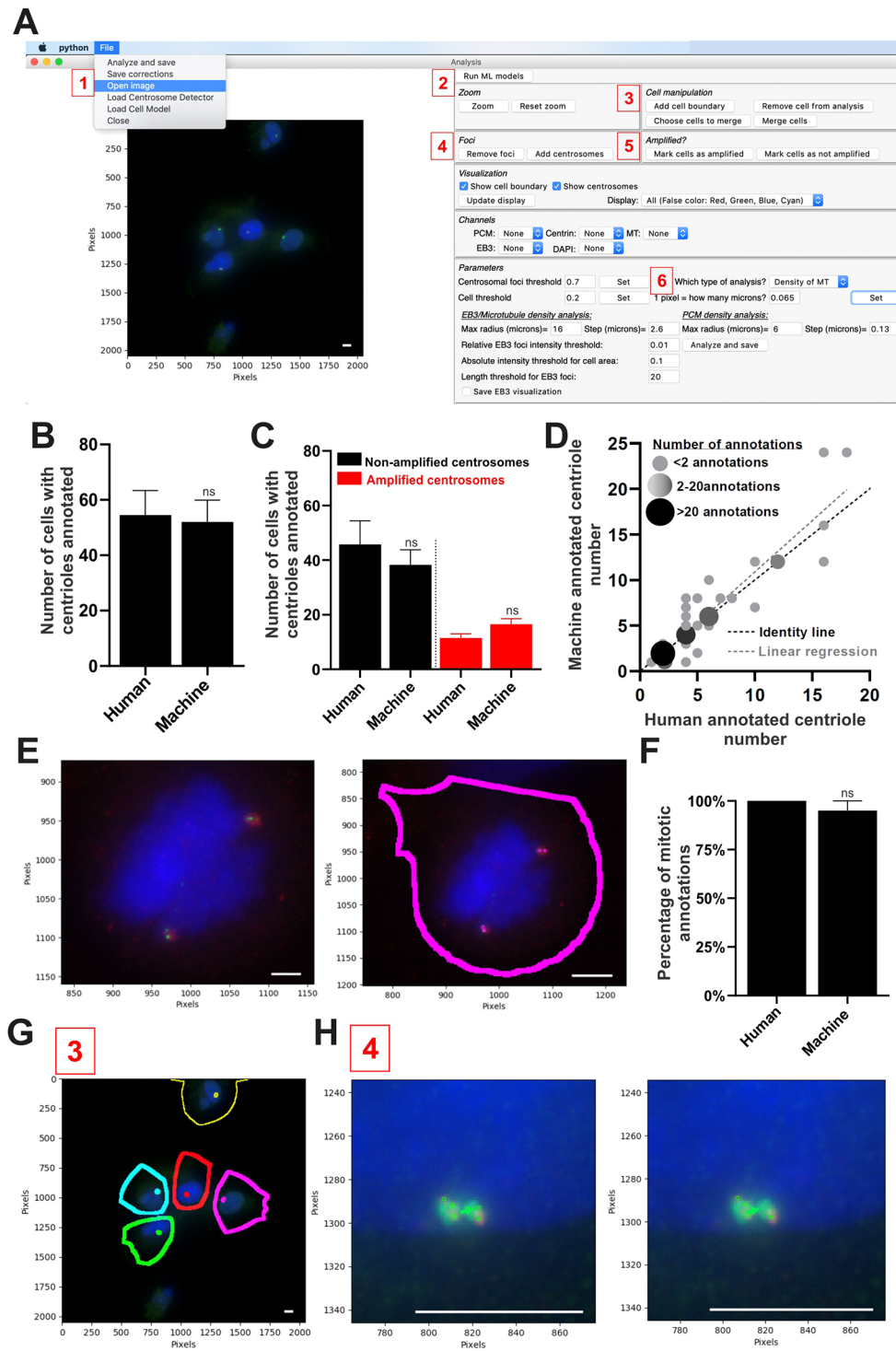


Fig. 3. A semi-automatic graphical tool for performing PCM and MT density analysis. (A) Screenshot of the graphical interface that encapsulates the machine learning-aided detection of centrosomes with image processing algorithms to analyze the spatial distribution of MT organization. The steps to perform the PCM and MT density analysis are shown in red boxes. Scale bar: 10 μ m.

(B) Comparison between the number of cells with centrosomes detected by machine and the number detected by human annotation. Data are shown as mean \pm s.e.m. of four independent replicates. (C) Comparison between number of centrosome-amplified (red) and non-amplified (black) cells with centrosomes detected by machine or human. Data are shown as mean \pm s.e.m. of $n \geq 3$ replicates. (D) Distribution of centriole numbers detected by machine versus humans. Each dot represents the number of centrioles annotated in each cell by the machine and the human. The dot size represents the number of cells that have the same pair of values: bigger black circles represent a higher number of annotations while smaller gray circles represent a lower number of annotations. A total of 68 cells are represented. Only cells that were annotated by both the algorithm and human were considered. (E) Left panel, screenshot of the image in the graphical interface showing a mitotic cell stained for centrosomes (centrin, green), PCM (pericentrin, red), and chromosomes (DAPI, blue). Right panel, screenshot of the image in the graphical interface showing detected centrosomes (magenta) in the mitotic cell. Scale bars: 10 μ m. (F) Percentage of mitotic centrosomal annotations detected by machine versus human. Data are shown as mean \pm s.e.m. percentages of mitotic annotations by human and by machine, computed over two independent replicates. (G) Screenshot of an image in the graphical interface showing edited cell boundaries (red, green, magenta, white, cyan, and yellow boundaries). Scale bar: 10 μ m. (H) Left panel, screenshot of a spurious detection (green arrow) as a centrosome. Right panel, screenshot after removal of the spurious detection (green arrow) in the graphical interface. Circles indicate features annotated as centrosomes by the detection algorithm. Scale bars: 10 μ m. ns, not significant (Wilcoxon test).

centrosome relative to that at a non-amplified centrosome (Fig. S2A,B). This suggests that γ -tubulin is not only elevated in the entire cluster of amplified centrosomes but also elevated per centrosome in an amplified centrosome. Furthermore, the relative γ -tubulin intensity diminished by at least 70% outside the pericentriolar space (between 1.0 to 2.0 μ m), similar to the intensity profile seen for non-amplified centrosomes (Fig. 4B). This suggests that while the PCM protein levels are elevated at the core centrosome, γ -tubulin remains constrained to the pericentriolar space of the amplified centrosomes. Consistent with the increase in

centrosomal γ -tubulin intensity, pericentrin protein was also elevated at centrosomes in breast cancer cells with amplified centrosomes (Fig. 4C,D). The magnitude of increase in both γ -tubulin and pericentrin intensity correlated with increasing numbers of centrioles. Intensities gradually and continuously increased up to two-fold relative to the levels in cells with non-amplified centrosomes (Fig. S2C,D). However, unlike γ -tubulin, the pericentrin protein intensity at the amplified centrosomes did not diminish as rapidly outside the core pericentriolar space when compared to non-amplified centrosomes.

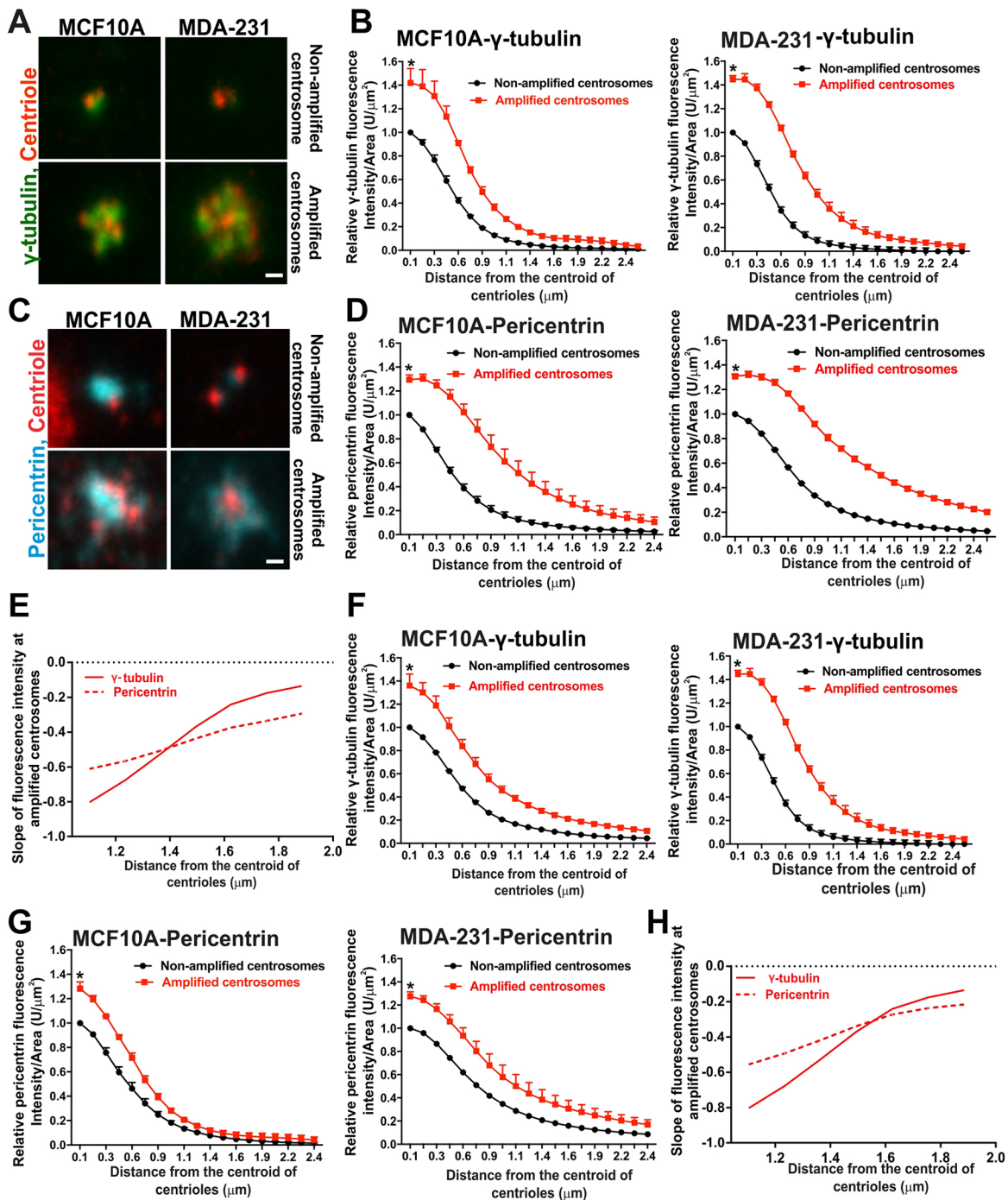


Fig. 4. Centrosome-amplified breast cancer cells exhibit PCM defects. (A) Non-amplified and amplified centrosomes in MCF10A cells (left) and MDA-231 cells (right) stained for centrioles (centrin, red) and γ -tubulin (green). Scale bar: 1 μ m. (B) Left panel, γ -tubulin fluorescence intensity per unit area in amplified (red) cells relative to non-amplified (black) MCF10A cells obtained using human annotation ($P=1.9\times 10^{-6}$). Right panel, γ -tubulin fluorescence intensity per unit area in amplified (red) cells relative to non-amplified (black) MDA-231 cells obtained using human annotation ($P=1.9\times 10^{-6}$). (C) Non-amplified and amplified centrosomes in MCF10A cells (left) and MDA-231 cells (right) stained for centrioles (centrin, red) and pericentrin (cyan). Scale bar: 1 μ m. (D) Left panel, pericentrin fluorescence intensity per unit area in amplified (red) cells relative to non-amplified (black) MCF10A cells obtained using human annotation ($P=3.8\times 10^{-6}$). Right panel, pericentrin fluorescence intensity per unit area in amplified (red) cells relative to non-amplified (black) MDA-231 cells obtained using human annotation ($P=1.9\times 10^{-6}$). (E) The slope, or first derivative, of fluorescence intensities of γ -tubulin and pericentrin at amplified centrosomes outside the core pericentriolar space (between 1.0 μ m and 2.0 μ m). (F) Left panel, γ -tubulin fluorescence intensity per unit area in amplified (red) cells relative to non-amplified (black) MCF10A cells obtained using machine annotation ($P=3.8\times 10^{-6}$). Right panel, γ -tubulin fluorescence intensity per unit area in amplified (red) cells relative to non-amplified (black) MDA-231 cells obtained using machine annotation ($P=3.8\times 10^{-6}$). (G) Left panel, pericentrin fluorescence intensity per unit area in amplified (red) cells relative to non-amplified (black) MCF10A cells obtained using machine annotation ($P=3.8\times 10^{-6}$). Right panel, pericentrin fluorescence intensity per unit area in amplified (red) cells relative to non-amplified (black) MDA-231 cells obtained using machine annotation ($P=3.8\times 10^{-6}$). (H) The slope, or first derivative, of fluorescence intensities of γ -tubulin and pericentrin at amplified centrosomes outside the core pericentriolar space (between 1.0 μ m and 2.0 μ m). In B,D,F,G, data are shown as mean+s.e.m., see Materials and Methods for numbers of cells analyzed. U μ m⁻², units μ m⁻². * $P<0.05$. P -values were calculated using the Wilcoxon test.

To quantify the differences in how γ -tubulin and pericentrin diminish outside the pericentriolar cloud, we estimated the first derivative, or slope, of γ -tubulin and pericentrin fluorescence intensities outside the core pericentriolar space (Fig. 4E; Fig. S2E). The slope of γ -tubulin intensity at amplified centrosomes was more negative in comparison to the slope of pericentrin intensity at amplified centrosomes. The γ -tubulin fluorescence intensity diminished quickly outside the core pericentriolar space at amplified centrosomes in comparison to pericentrin fluorescence intensity (Fig. 4E; Fig. S2E). Moreover, this suggests that the increase in centrioles at amplified centrosomes modulates the organization of specific pericentriolar proteins.

We next utilized the semi-automatic machine learning algorithm to quantify the pericentriolar defects at amplified centrosomes that were described using manual annotation and quantification above. The semi-automatic graphical tool annotated cells with greater than two centrosomes as amplified. However, if the cells were mis-annotated as non-amplified, then the annotations were manually edited to amplified centrosomes. The tool corroborated the results from the manual quantification: higher γ -tubulin levels in amplified relative to non-amplified centrosomes were reported by the semi-automated analysis (Fig. 4B,F). Although the centrosomal γ -tubulin intensity was elevated, the distribution remained encompassed within the pericentriolar space (60% of total fluorescence intensity) of the amplified centrosomes. Pericentrin was also elevated at centrosomes in breast cancer cells with CA (Fig. 4D,G). Furthermore, pericentrin expanded the pericentriolar space of amplified centrosomes (Fig. 4H). These results show that the semi-automatic machine learning algorithm detects the same centrosome organization defects at amplified centrosomes as found by the more laborious manual annotation and quantification.

MT density is increased at amplified centrosomes

The increase in γ -tubulin at amplified centrosomes suggests that MTs might also be elevated around centrosomes. To test whether cells with amplified centrosomes have increased MTs, we measured the MT fluorescence intensity per unit area using the algorithm. MT intensity was 30% greater in both MCF10A and MDA-231 cells with amplified centrosomes compared to cells without CA (Fig. 5A,B). Overall, these data suggest that centrosome-amplified breast cancer cells have an increase in centrosome-associated γ -tubulin that promotes MT nucleation to increase the MT density.

To test whether amplified centrosomes increase the number of growing MT ends, we created stably expressing tetracycline-inducible mNeon-EB3-expressing MCF10A and MDA-231 cell lines. Tetracycline was added for 48 h to induce expression of mNeon-EB3, and the number of EB3 comets per unit area was measured using the EB3 density feature in the graphical interface (Fig. 5C-H). A 20% and 40% increase in the EB3 density was observed at amplified centrosomes relative to the non-amplified centrosomes in MCF10A and MDA-231 cells, respectively. This suggests that amplified centrosomes nucleate more MT ends compared to non-amplified centrosomes. Furthermore, the density of EB3 foci was elevated closer to centrosomes in comparison to density at the cell periphery in both MCF10A and MDA-231 cells (Fig. 5E,H). This is consistent with the elevated MT density that is observed closer to centrosomes relative to the cell periphery (Fig. 5B). Taken together, the data suggest that MT nucleation is enriched at amplified centrosomes in comparison to that at non-amplified centrosomes of MCF10A and MDA-231 cells.

DISCUSSION

A semi-automatic graphical tool for centrosome detection and analysis of MT organization

We designed a semi-automatic graphical tool that can speed up the analysis of centrosomes and MTs. It does so by performing an initial automated detection of centrosomes that can later be corrected, and it removes the need for manual identification of EB3 comets. Because of limitations in the amount of training the machine learning algorithm can realistically incorporate, we found that user correction of the algorithm's predictions was crucial to ensure the accuracy of the analysis. These corrections can be used to further train the underlying model, allowing the model to improve with use. However, in our experiments, we did not perform such retraining so as to keep the model fixed during analysis. Furthermore, retraining provides flexibility to the tool, allowing it to be used for different experiments. Training scripts and models are publicly available to the research community (<https://bharath272.github.io/centrosome-analysis/>).

While correcting the output from the algorithm, we found several common modes of error. First, unlike internet images where machine-learning technologies are routinely used, the dynamic range of fluorescence images can vary significantly, and naïve training of convolutional networks does not generalize across such variations. We addressed this by normalizing each image independently to have zero mean and unit variance. Second, we found that while the algorithm correctly identified the rough location of the centrosomes, it often did not correctly estimate the number of centrioles. We thus had to correct its predictions as to which cells exhibited CA. Accurate counting has not been addressed in the computer vision literature and deserves further attention (Chattopadhyay et al., 2017). Finally, we found that the algorithm often merged nearby cells or estimated their boundary incorrectly. These issues reveal challenges that must be addressed by future computer vision research.

In this study, we have used machine learning for the segmentation of cells and the detection of centrosomes. We used machine learning for these two steps of the analysis because these are the steps that are difficult to specify by hand as precise computational procedures. Other steps of the analysis, such as the measurement of the fluorescence intensity or the computation of the centroid of the centrosomes, are precise computations that can be created without resorting to machine learning. Nevertheless, our results show that machine learning techniques can provide a significant degree of automation and partly mitigate the difficulties associated with manual annotation.

PCM proteins and their organization around amplified centrosomes in breast cancer

Taxol (paclitaxel), Taxotere, and cabazitaxel are MT-stabilizing drugs that have been used in chemotherapy for the treatment of cancers (Ho and Mackey, 2014; Pazdur et al., 1993; Schiff and Horwitz, 2006). These drugs promote mitotic arrest. However, the proportion of cells lost by mitotic arrest does not explain the efficacy of Taxol as a chemotherapy agent. The alternate hypothesis that Taxol affects interphase breast cancer cells has not been well studied (Weaver, 2014). Specifically, it is unclear whether these MT-stabilizing drugs differentially impact underlying differences in the interphase centrosomes and MTs of normal compared to transformed breast cells (Weaver, 2014). To understand the differences in interphase centrosomes and MTs of breast cancer cells, we investigated the levels and the distribution of centrosome PCM proteins and MTs using our semi-automatic machine learning-aided approach. We found that amplified centrosomes in interphase breast cancer cells not only have elevated γ -tubulin and pericentrin but that the distribution of pericentrin localization is expanded. This

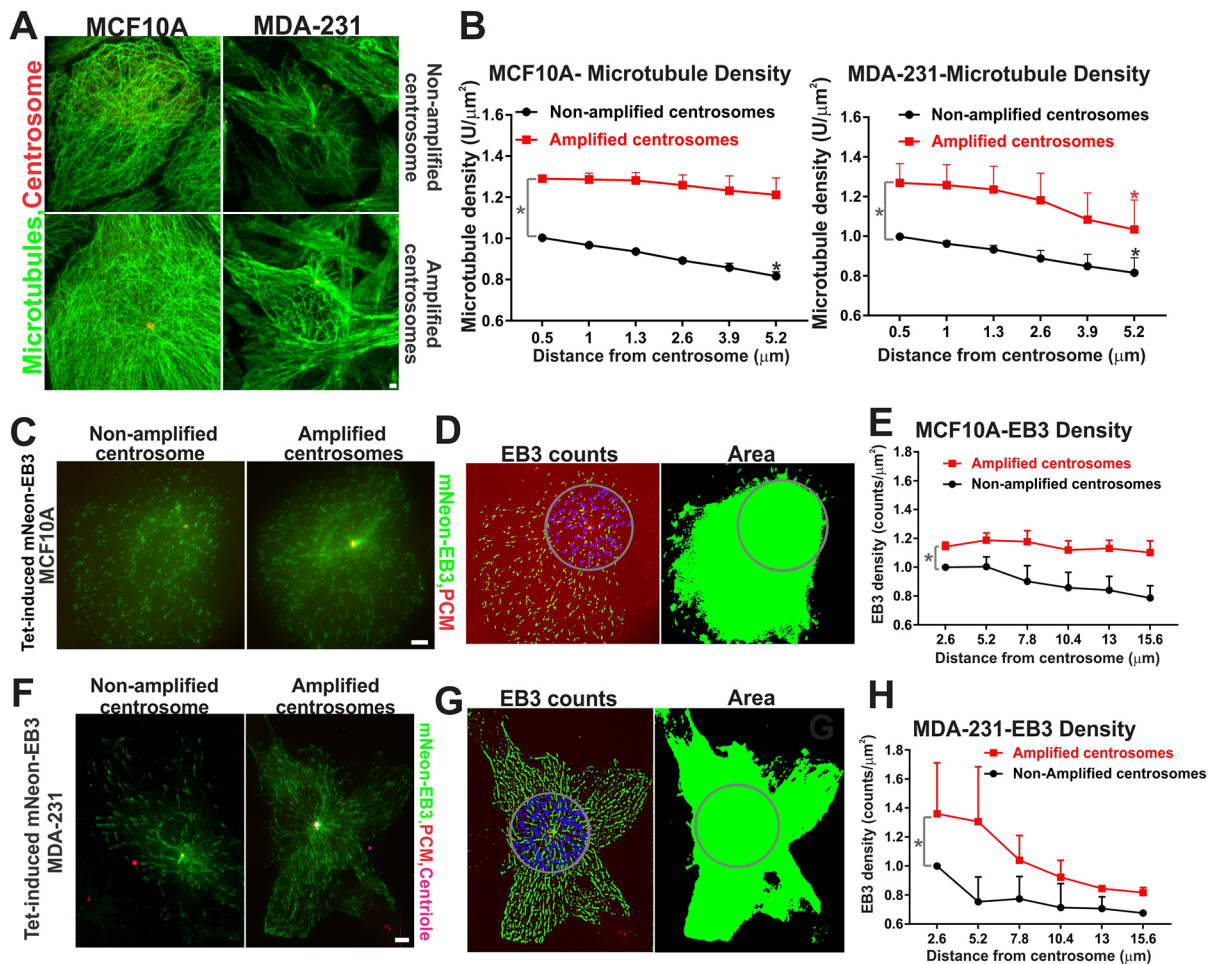


Fig. 5. Centrosome-amplified breast cancer cells exhibit increased MT density. (A) Non-amplified and amplified centrosomes in MCF10A cells (left) and MDA-231 cells (right) stained for centrioles (CEP192, red) and MTs (α -tubulin, green). Scale bar: 10 μ m. (B) Left panel, MT fluorescence intensity in amplified (red) MCF10A cells relative to non-amplified (black) cells (P -values: gray asterisk, $P=4 \times 10^{-6}$; black asterisk, $P=3.4 \times 10^{-6}$). Right panel, MT fluorescence intensity in amplified (red) MDA-231 cells relative to non-amplified (black) cells (P -values: gray asterisk, $P=8 \times 10^{-6}$; black asterisk, $P=2 \times 10^{-2}$; red asterisk, $P=9.3 \times 10^{-3}$). (C) Non-amplified and amplified centrosomes in Tet-induced mNeon-EB3 MCF10A cells stained for centrioles (centrin, green) and γ -tubulin (red). Scale bar: 5 μ m. (D) Processed images of non-amplified Tet-induced mNeon-EB3 MCF10A cells utilized for the measurement of EB3 counts and area. Circle indicates region of interest. (E) EB3 foci counts per unit area in non-amplified (black) and amplified (red) MCF10A cells (gray asterisk, $P=3.1 \times 10^{-2}$). (F) Non-amplified and amplified centrosomes in Tet-induced mNeon-EB3 MDA-231 cells stained for centrioles (centrin, green) and γ -tubulin (red). Scale bar: 5 μ m. (G) Processed images of amplified Tet-induced mNeon-EB3 MDA-231 cells utilized for the measurement of EB3 counts and area. Circle indicates region of interest. (H) EB3 counts per unit area in amplified (red) MDA-231 cells relative to non-amplified (black) cells (gray asterisk, $P=3.1 \times 10^{-2}$). Data in B,E,H are shown as mean \pm s.e.m. See Materials and Methods for numbers of cells analyzed. $U \mu m^{-2}$, units μm^{-2} . Gray asterisks correspond to tests for significant differences in mean at each distance from the centrosome (paired Student's t -test). This tests whether the red dots are consistently higher than the black dots. Black asterisks correspond to tests for significant differences in mean non-amplified microtubule intensity between 0.5 μ m and 5.2 μ m (Wilcoxon test). Red asterisk corresponds to a test of significant difference in mean amplified microtubule intensity between 0.5 μ m and 5.2 μ m (Wilcoxon test).

suggests that there are more PCM sites that promote MT nucleation and MTs that are organized around amplified centrosomes. This might disrupt MT-related processes like cellular motility, ciliogenesis, signaling, and MT-dependent transport (Caviston and Holzbaur, 2006; Desai and Mitchison, 1997; Hyman and Karsenti, 1996; Inoue and Salmon, 1995). Furthermore, expansion of the PCM may interfere with MT-motor-dependent trafficking (Galati et al., 2018; Nanjundappa et al., 2019). This altered MT-motor-dependent trafficking may affect responses to MT-stabilizing drugs.

The interphase MT network is altered in centrosome-amplified breast cancer cells

In addition to the increased PCM, we found that the density of MTs was elevated in MCF10A and MDA-231 cells with amplified

centrosomes compared to those with non-amplified centrosomes. This increased MT organization could be attributed to cell cycle differences, because S-phase cells have elevated microtubule density relative to G1-phase cells (Salaycik, 2005). However, the pool of non-amplified centrosomes included interphase cells with either one (likely G1-phase) or two (likely S-phase) centrosomes. Hence, cell cycle differences likely do not explain the increase in MT organization observed in the centrosome-amplified cell population. An increase in microtubule nucleation from additional centrosome-proximal PCM sites is a more likely explanation for the increase in MT density that was observed with centrosome amplification. Finally, whether cytoplasmic α/β -tubulin concentrations are elevated in these cells to promote increased MT assembly is not known. Nevertheless, our studies suggest that amplified centrosomes promote an increased MT network in interphase breast cancer cells.

Similar to MT density, we observed that the number of growing MTs, as judged by EB3 density, is elevated in centrosome-amplified breast cancer cells relative to that in cells without centrosome amplification (Fig. 5E,H). Furthermore, our results revealed differences in MT populations between MDA-231 and MCF10A cells. EB3 density decreased towards the cell periphery of centrosome-amplified MDA-231 cells relative to the density in centrosome-amplified MCF10A cells. MT density also decreased towards the cell periphery in centrosome-amplified MDA-231 cells (Fig. 5B). However, unlike MDA-231 cells, MT density in MCF10A cells with amplified centrosomes was greater at the cell periphery (Fig. 5B). The increased MT density from MCF10A amplified centrosomes at the cell periphery might be explained either by the presence of longer and more stable MTs or by increased nucleation and branching of the MT network (Goshima et al., 2008; Ishihara et al., 2014; Petry et al., 2013). It remains unclear why MCF10A and MDA-231 cells exhibited these unique profiles in MT growth and density. An increase in MT density and EB3 comets around amplified centrosomes reflects differences in MTs between non-amplified and amplified interphase breast cancer cells. Such altered cytoskeletal architecture could lead to changes in intracellular trafficking and disrupt cellular processes including ciliogenesis, cell migration and cell polarization (Bouchet and Akhmanova, 2017; Caviston and Holzbaur, 2006; Siegrist and Doe, 2007).

In summary, we built a semi-automated tool with a graphical interface that enables quantitative measurements of centrosome aberrations. Using this approach, we detected centrosome aberrations in cancer cells and showed an increase in MT nucleation that promotes changes to MT density and the spatial distribution of MTs around centrosomes. This not only highlights how the use of the machine learning-based approaches for the detection of centrosomal aberrations can speed up manual analysis but also reveals how such semi-automatic analysis can be globally applied to quantitative cell biological problems.

MATERIALS AND METHODS

Cell culture

Breast cancer cell lines MCF10A and MDA-MB-231 (MDA-231) were obtained from the University of Colorado Cancer Center Tissue Culture Core. Mammalian tissue culture lines were all grown (following a method similar to that described by Ganapathi Sankaran et al., 2019) at 37°C with 5% CO₂. MCF10A cells were received at passage 51 and were grown in DMEM/F12 (Invitrogen #11330-032), 5% horse serum (Invitrogen #16050-122), 20 ng ml⁻¹ EGF (Invitrogen #PHG0311), 0.5 mg ml⁻¹ hydrocortisone (Sigma #H-0888), 100 ng ml⁻¹ cholera toxin (Sigma #C-8052), 10 µg ml⁻¹ insulin (Sigma #I-1882) and 1% penicillin and streptomycin (Pen/Strep; Invitrogen #15070-063). MDA-MB-231 cells, received at passage 15, were grown in DMEM (Invitrogen #11965-092), Pen/Strep (Invitrogen #15070-063) and 10% FBS (FBS; Gemini Biosciences). Cell lines were authenticated at the sources and tested negative for mycoplasma using the MycoAlert mycoplasma detection kit through the University of Colorado Cancer Center Tissue Culture Core. Cells were passaged and sub-cultured using trypsin (Invitrogen #150901-046) when cultures reached 60–80% confluency (Ganapathi Sankaran et al., 2019).

Generation of tetracycline-inducible mNeon-EB3 MCF10A and MDA-231 cells

The EB3–mNeon (C-terminal fusion) fragment was obtained through PCR with Phusion DNA polymerase of a pre-existing plasmid (pmNeonGreen-EB3-7; Allele Biotechnology) using primers that have NheI and XmaI sites appended to them. This was cloned into the tetracycline-inducible construct pcw57.1 using the enzymes NheI and AgeI. The resulting C-terminal fusion construct is referred to as tetracycline-inducible mNeon–EB3 in the text.

Lentivirus harboring tetracycline-inducible mNeon–EB3 were made by the transfection of 293FT cells. 293FT cells were plated in 6-cm dishes and allowed to reach 50%–70% confluency. Cells were then transfected with tetracycline-inducible EB3–mNeon constructs, and second-generation lentivirus packaging plasmids (pMD2.G and psPAX2) using Lipofectamine 2000 (Life Technologies # 11668019). The 293FT cells, the viral constructs pMD2.G and psPAX2, and pcw57.1 were obtained from Dr Heide Ford, University of Colorado, Aurora, CO. 293FT medium containing virus was harvested and MDA-231 and MCF10A cells were infected for 24–48 h in the presence of 10 µg ml⁻¹ (26.7 µM) polybrene. After a 24 h recovery, transduced cells were selected with puromycin at 2 µg ml⁻¹ (4.24 µM) and were flow-sorted to isolate and plate single cells into 96-well plates. Such clones were cultured in 50% filtered conditioned medium with 50% fresh medium. Tetracycline-inducible mNeon–EB3 MCF10A and MDA-231 cells were induced with tetracycline (Invitrogen #550205) at 2.5 µg ml⁻¹ (5.63 µM) (Ganapathi Sankaran et al., 2019).

Transfections

293FT cells at 50–80% confluence were transfected using Lipofectamine 2000 (Invitrogen # 11668019). Plasmid DNA and Plus reagent (Invitrogen # 11514015) were mixed at 1:1 and incubated for 5 min. This mixture was then combined with Lipofectamine at a 1:3 ratio. Complexes were diluted in Opti-MEM (Invitrogen # 31985062). After a 4 h incubation, the complexes were removed and the transfected cells were supplied with fresh medium (Ganapathi Sankaran et al., 2019).

Immunofluorescence

12-mm diameter coverslips were acid-washed and heated to 50°C in 100 mM HCl for 16 h. This was followed by washes with water, 50%, 70%, and 95% ethanol for 30 min each. Coverslips were coated with type-1 collagen (Sigma # C9791), air-dried for 20 min in the laminar hood and exposed to UV light for cross-linking of collagen for 20 min (procedure similar to that described in Ganapathi Sankaran et al., 2019). Cells were cultured on collagen-coated coverslips to 55–70% confluence. For centrosome immunofluorescence, cells were fixed with 100% methanol at –20°C for 8 min. Fixed cells were washed with PBS/Mg (1× PBS and 1 mM MgCl₂), and then blocked with Knudsen Buffer (1× PBS, 0.5% BSA, 0.5% NP-40, 1 mM MgCl₂ and 1 mM NaN₃) for 1 h. Cells were incubated overnight with primary antibodies diluted in Knudsen Buffer at 4°C. Coverslips were washed with PBS three times in 5-min intervals. Secondary antibodies and Hoechst 33258 (10 µg ml⁻¹, Sigma #B2261) were diluted in Knudsen buffer and incubated for 1 h at room temperature. Coverslips were mounted using Citifluor (Ted Pella) and sealed with clear nail polish (Ganapathi Sankaran et al., 2019). Antibodies used for immunofluorescence were anti-centrin (1:2000; 20H5; Abcam), anti-pericentrin (1:2000; Abcam), anti-CEP192 (1:2000; obtained from Dr Andrew Holland, Johns Hopkins University, Baltimore, MD), anti-γ-tubulin (1:1000; DQ-19; Sigma) and anti-α-tubulin (1:500; DM1A; Sigma). Alexa-Fluor secondary antibodies were diluted to 1:1000 for all experiments (Molecular Probes).

S-phase arrested and M-phase cells

Cells were grown on collagen-coated coverslips overnight using the procedure described above. Cells were treated with 1.6 µg ml⁻¹ aphidicolin for 22 h and then fixed with methanol. For the M-phase cells, cells were arrested in S-phase as above for 21 h, washed three times with medium to remove the drug and allowed to progress out of the arrest for 10 h. They were then fixed with methanol using the procedure described above.

Microscopy

The fluorescence imaging, as shown in the figures, used methods identical to those described in Dahl et al. (2015). Briefly, images were acquired using a Nikon TIE (Nikon Instruments, Inc.) inverted microscope stand equipped with a 100× PlanApo DIC, NA 1.4 objective. Images were captured using an Andor iXon EMCCD 888E camera or an Andor Xyla 4.2 CMOS camera (Andor Technologies). Images in Fig. 4A,B were acquired using a Swept Field Confocal system (Prairie Technologies/Nikon Instruments) on a Nikon

Ti inverted microscope stand equipped with a 100× Plan Apo λ, NA 1.45 objective. Images were captured with an Andor Clara CCD camera (Andor Technologies) (Ganapathi Sankaran et al., 2019).

Nikon NIS Elements imaging software was used for image acquisition. Image acquisition times were constant within a given experiment and ranged from 50 to 400 ms, depending on the experiment. All images were acquired at approximately 25°C. We utilized maximum intensity projections of the complete *z*-stacks for visualization of centrosomes. Maximum intensity projections of centrosomes compromise spatial information but have advantages over sum projections, which oversample centrosomes and collect more out of focus light, thus exaggerating differences in fluorescence intensity during quantification. The findings of previous studies support the visualization of centrosomes using maximum intensity projections (Dahl et al., 2015; Ganapathi Sankaran et al., 2020 preprint; Holland et al., 2010). For these reasons, maximum intensity projections were utilized. Adjustments to images presented in the figures involved only global, linear intensity changes. In the schematics (Fig. 1A,F; Fig. S1), for visualization of the neural network outputs, which can have an arbitrary dynamic range, global non-linear intensity normalization was performed to remove saturation artifacts and enhance contrast. This was done mainly to visualize the neural network outputs and does not add, remove, or alter any feature.

Centriole and centrosome number counts

Cells were scored as amplified, non-amplified and under duplicated based on centrin and γ -tubulin staining (Dahl et al., 2015). We classified cells with one or two centrosomes as non-amplified and cells with greater than two centrosomes as amplified. In other words, cells with two centrin and one γ -tubulin foci (one centrosome) or four centrin and two γ -tubulin foci (two centrosomes) were classified as non-amplified, whereas cells with greater than four centrin and two γ -tubulin foci were classified as amplified. Cells with three centrin foci and one γ -tubulin foci were classified as amplified, whereas cells with three centrin foci and two γ -tubulin foci were classified as under duplicated.

Computational tools

The semi-automatic graphical user interface was created as a standalone Python program. The training script is a separate program packaged with the graphical interface. The version of Python used was 3.6, and it was used in conjunction with Anaconda as a package manager. The graphical user interface accepts TIFF files. The size that a pixel denotes can be input through the interface. The graphical user interface was created using the Tkinter Python library. The underlying machine learning models were built using the PyTorch library; we used version 0.4.1. The code also relies on the NumPy library for matrix computations, the SciPy library for a variety of signal processing routines, the Matplotlib library for plotting and the TiffFile library for reading and writing TIFF images. The training annotations were collected using a separate stand-alone program with a graphical user interface.

The centrosome detection model was trained using stochastic gradient descent with momentum (Krizhevsky et al., 2012). The parameters of the training procedure were as follows: learning rate, 0.01; momentum, 0.9; weight decay, 0.0001; the total number of epochs, 100; batch size, 1. Because pixels on centrosomes are very few in number, the learning problem was found to be extremely imbalanced. To deal with this imbalance, the training loss on centrosomal pixels was increased by a factor of 1000. Such scaling is common in the machine learning literature whenever the dataset is imbalanced (Cui et al., 2019). The scale is usually chosen based on empirical accuracy of the model on the training set. This is mathematically equivalent to statistically sampling centrosomal pixels 1000 times more often during the stochastic gradient descent procedure. Such scaling mitigates the problem of an imbalanced dataset: without it, the model fails to detect any centrosomes. However, the scaling is merely a heuristic and does not completely eliminate the problem of dataset imbalance. Pixels within five pixels of a centrosome were ignored during training to avoid penalizing the algorithm for small deviations.

The cell segmentation model was also trained using stochastic gradient descent with momentum. The parameters of the training procedure were as follows: learning rate, 0.01; momentum, 0.9; weight decay, 0.0001; the total

number of epochs, 100; batch size, 1. The tolerance for the random walk segmenter was set to 0.01.

The centrosome analysis code is available here: <https://bharath272.github.io/centrosome-analysis/>. Also included with the code is a training script that can be used to train a new model based on the saved corrections from the graphical tool. This training script can be run from the terminal as follows: `python train_foci_detector.py --trainfiles trainingset.csv --modelfile foci_model.pt`.

Here, `trainingset.csv` must be a CSV (comma-separated values) file formatted as follows. Each row of this file consists of the path to the image (a TIFF file), and the path to the corresponding corrected annotation (a JSON file, saved from the graphical tool using the ‘Save corrections’ feature), separated by a comma. Such a CSV can be created using most spreadsheet software including Microsoft Excel. Once run, the training script will save the resulting trained foci detector in the file `foci_model.pt`.

Statistics and biological replicates

All center values represent means and error bars represent the standard error of the mean. All the experiments in the figures were performed using at least three independent biological replicates, except for Fig. 3E which was done in two independent biological replicates. The number of cells used for training the algorithm were as follows: Fig. 1C, 40 cells; Fig. 1D, 40 cells; Fig. 1E, 40 cells. The number of cells used in each immunofluorescence experiment were as follows: Fig. 3B, 52 cells per experiment (215 cells in total, four independent replicates); Fig. 3C, 40 cells per experiment; Fig. 3D, 40 cells per experiment; Fig. 3F, 12 mitotic cells (two independent replicates); Fig. 4A,B,F, 40 cells per condition/80 cells per experiment (~40 centrosomes per condition); Fig. 4C,D,G, 40 cells per condition/80 cells per experiment (~40 centrosomes per condition); Fig. 5A,B, 30 cells per condition/60 cells experiment; Fig. 5C–E, 40 cells per experiment; Fig. 5F–H, 40 cells per experiment.

Fisher’s test was utilized to examine the significance of contingency when data were classified into two or more categories. A Student’s two-tailed unpaired *t*-test was used to examine significance between two normal unpaired distributions (equal variance assumed). Student’s two-tailed paired tests were used to examine significance between two normal continuous paired distributions. Pairing was assumed when two measurements were taken for an individual variable. For example, in Fig. 5B every measurement corresponding to non-amplified and amplified centrosomes, at a particular distance from the centrosome (0.5, 1, 1.3, 2.6, 3.9 and 5.2 μ m), were considered ‘paired’. Normality tests were performed both on the raw data and meta-data extracted from the replicates of raw data. The Shapiro–Wilk normality test and D’Agostino–Pearson omnibus normality test were utilized to examine the normality of data. The Shapiro–Wilk normality test was used when the number of samples was less than eight. When the number of samples was greater than eight, the D’Agostino–Pearson omnibus normality test was used. The Mann–Whitney *U*-test was utilized to examine the significance of non-normal unpaired distributions, and the Wilcoxon test was used to examine the significance of non-normal paired distributions. Results were considered statistically significant with $P < 0.05$. *P*-values are denoted in the figure legends.

Acknowledgements

We thank Dr Heide Ford for viral constructs, Dr Andrew Holland for α -CEP192 antibodies, and UC Tissue culture core for the cell lines.

Competing interests

The authors declare no competing or financial interests.

Author contributions

Conceptualization: D.G.S., C.G.P.; Methodology: D.G.S., B.H., C.G.P.; Software: B.H.; Validation: D.G.S., B.H.; Formal analysis: D.G.S., B.H.; Investigation: D.G.S.; Resources: B.H.; Data curation: D.G.S., B.H.; Writing - original draft: D.G.S., B.H.; Writing - review & editing: A.J.S.-W., B.L.M., B.H., C.G.P.; Visualization: A.J.S., B.L.M.; Supervision: B.H., C.G.P.; Project administration: C.G.P.; Funding acquisition: C.G.P.

Funding

C.G.P. is supported by the American Cancer Society (RSG-16-157-01-CCG), the National Institutes of Health National Institute of General Medical Sciences

(GM099820) and the Boettcher Foundation. Deposited in PMC for release after 12 months.

Supplementary information

Supplementary information available online at <https://jcs.biologists.org/lookup/doi/10.1242/jcs.243543.supplemental>

Peer review history

The peer review history is available online at <https://jcs.biologists.org/lookup/doi/10.1242/jcs.243543.viewer-comments.pdf>

References

- Allen, C. and Borisy, G. G. (1974). Structural polarity and directional growth of microtubules of *Chlamydomonas flagella*. *J. Mol. Biol.* **90**, 381-402. doi:10.1016/0022-2836(74)90381-7
- Applegate, K. T., Besson, S., Matov, A., Bagonis, M., Jaqaman, K. and Danuser, G. (2011). plusTipTracker: quantitative image analysis software for the measurement of microtubule dynamics. *J. Struct. Biol.* **176**, 168-184. doi:10.1016/j.jsb.2011.07.009
- Bergen, L. G. and Borisy, G. G. (1980). Head-to-tail polymerization of microtubules in vitro. *J. Cell Biol.* **84**, 141-150. doi:10.1083/jcb.84.1.141
- Bouchet, B. P. and Akhmanova, A. (2017). Microtubules in 3D cell motility. *J. Cell Sci.* **130**, 39-50. doi:10.1242/jcs.189431
- Brinkley, B. R. (1985). Microtubule organizing centers. *Annu. Rev. Cell Biol.* **1**, 145-172. doi:10.1146/annurev.cb.01.10.145
- Caviston, J. P. and Holzbaur, E. L. F. (2006). Microtubule motors at the intersection of trafficking and transport. *Trends Cell Biol.* **16**, 530-537. doi:10.1016/j.tcb.2006.08.002
- Chattopadhyay, P., Vedantam, R., Selvaraju, R. R., Batra, D. and Parikh, D. (2017). Counting everyday objects in everyday scenes. *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017 2017-Janua*, 4428-4437. doi:10.1109/CVPR.2017.471
- Cui, Y., Jia, M., Lin, T. Y., Song, Y. and Belongie, S. (2019). Class-balanced loss based on effective number of samples. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* **2019-June**, 9260-9269. doi:10.1109/CVPR.2019.00949
- D'Assoro, A. B., Barrett, S. L., Folk, C., Negron, V. C., Boeneman, K., Busby, R., Whitehead, C., Stivala, F., Lingle, W. L. and Salisbury, J. L. (2002). Amplified centrosomes in breast cancer: a potential indicator of tumor aggressiveness. *Breast Cancer Res. Treat.* **75**, 25-34. doi:10.1023/A:1016550619925
- Dahl, K. D., Sankaran, D. G., Bayless, B. A., Pinter, M. E., Galati, D. F., Heasley, L. R., Giddings, T. H. and Pearson, C. G. (2015). A short CEP135 splice isoform controls centriole duplication. *Curr. Biol.* **25**, 2591-2596. doi:10.1016/j.cub.2015.08.039
- Denu, R. A., Zasadil, L. M., Kanugh, C., Laffin, J., Weaver, B. A. and Burkard, M. E. (2016). Centrosome amplification induces high grade features and is prognostic of worse outcomes in breast cancer. *BMC Cancer* **16**, 1-13. doi:10.1186/s12885-015-2026-y
- Desai, A. and Mitchison, T. J. (1997). Microtubule polymerization dynamics. *Annu. Rev. Cell Dev. Biol.* **13**, 83-117. doi:10.1146/annurev.cellbio.13.1.83
- Dicthenberg, J. B., Zimmerman, W., Sparks, C. A., Young, A., Carrington, W., Zheng, Y., Vidair, C., Fay, F. S. and Doxsey, S. J. (2002). Pericentriar and γ -tubulin form a protein complex and are organized into a novel lattice at the centrosome. *J. Cell Biol.* **141**, 163-174. doi:10.1083/jcb.141.1.163
- Farache, D., Emorine, L., Haren, L. and Merdes, A. (2018). Assembly and regulation of γ -tubulin complexes. *Open Biol.* **8**, 170266. doi:10.1098/rsob.170266
- Fong, K.-W., Choi, Y.-K., Rattner, J. B. and Qi, R. Z. (2007). CDK5RAP2 is a pericentriolar protein that functions in centrosomal attachment of the gamma-tubulin ring complex. *Mol. Biol. Cell* **19**, 308-317. doi:10.1091/mbc.e07-04-0371
- Fu, J. and Glover, D. M. (2012). Structured illumination of the interface between centriole and peri-centriolar material. *Open Biol.* **2**, 120104. doi:10.1098/rsob.120104
- Galati, D. F., Sullivan, K. D., Pham, A. T., Espinosa, J. M. and Pearson, C. G. (2018). Trisomy 21 represses cilia formation and function. *Dev. Cell* **46**, 641-650.e6. doi:10.1016/j.devcel.2018.07.008
- Ganapathi Sankaran, D., Stemm-Wolf, A. J. and Pearson, C. G. (2019). CEP135 isoform dysregulation promotes centrosome amplification in breast cancer cells. *Mol. Biol. Cell* **30**, 1129-1244. doi:10.1091/mbc.E18-10-0674
- Ganapathi Sankaran, D., Hariharan, B. and Pearson, C. G. (2020). A semi-automated machine learning-aided approach to quantitative analysis of centrosomes and microtubule organization. *bioRxiv* 2020.01.03.894071. doi:10.1101/2020.01.03.894071
- Girshick, R., Donahue, J., Darrell, T. and Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 580-587. doi:10.1109/CVPR.2014.81
- Goshima, G., Mayer, M., Zhang, N., Stuurman, N. and Vale, R. D. (2008). Augmin: A protein complex required for centrosome-independent microtubule generation within the spindle. *J. Cell Biol.* **181**, 421-429. doi:10.1083/jcb.2007111053
- Grady, L. (2006). Random walks for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**, 1768-1783. doi:10.1109/TPAMI.2006.233
- Guo, H., Gao, M., Ma, J., Xiao, T., Zhao, L., Gao, Y. and Pan, Q. (2007). Analysis of the cellular centrosome in fine-needle aspirations of the breast. *Breast Cancer Res.* **9**, R48. doi:10.1186/bcr1752
- He, K. and Sun, J. (2014). Convolutional neural networks at constrained time cost. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* **07-12-June**, 5353-5360.
- Hinchcliffe, E. H. and Sluder, G. (2001). "It takes two to tango": understanding how centrosome duplication is regulated throughout the cell cycle. *Genes Dev.* **15**, 1167-1181. doi:10.1101/gad.894001
- Ho, M. Y. and Mackey, J. R. (2014). Presentation and management of docetaxel-related adverse effects in patients with breast cancer. *Cancer Manag. Res.* **6**, 253-259. doi:10.2147/CMAR.S40601
- Holland, A. J., Lan, W., Niessen, S., Hoover, H. and Cleveland, D. W. (2010). Polo-like kinase 4 kinase activity limits centrosome overduplication by autoregulating its own stability. *J. Cell Biol.* **188**, 191-198. doi:10.1083/jcb.200911102
- Hyman, A. and Karsenti, E. (1996). Morphogenetic properties of microtubules and mitotic spindle assembly. *Cell* **84**, 401-410. doi:10.1016/S0092-8674(00)81285-4
- Inoue, S. and Salmon, E. D. (1995). Force generation by microtubule assembly/disassembly in mitosis and related movements shinya. *Mol. Biol. Cell* **6**, 1619-1640. doi:10.1091/mbc.6.12.1619
- Ishihara, K., Nguyen, P. A., Groen, A. C., Field, C. M. and Mitchison, T. J. (2014). Microtubule nucleation remote from centrosomes may explain how asters span large cells. *Proc. Natl. Acad. Sci. USA* **111**, 17715-17722. doi:10.1073/pnas.1418796111
- Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems 25* (ed. P. Bartlett), pp. 1097-1105. Red Hook, NY: Curran Associates, Inc.
- Lawo, S., Hasegan, M., Gupta, G. D. and Pelletier, L. (2012). Subdiffraction imaging of centrosomes reveals higher-order organizational features of pericentriolar material. *Nat. Cell Biol.* **14**, 1148-1158. doi:10.1038/ncb2591
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y. and Berg, A. C. (2016). SSD: Single shot multibox detector. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **9905 LNCS**, 21-37. doi:10.1007/978-3-319-46448-0_2
- Lopes, C. A. M., Mesquita, M., Cunha, A. I., Cardoso, J., Carapeta, S., Laranjeira, C., Pinto, A. E., Pereira-Leal, J. B., Dias-Pereira, A., Bettencourt-Dias, M. et al. (2018). Centrosome amplification arises before neoplasia and increases upon p53 loss in tumorigenesis. *J. Cell Biol.* **217**, 2353-2363. doi:10.1083/jcb.201711191
- Martel, G., Guerrero, A., Vieira, A. F., De Almeida, B. P., Machado, P., Mendonça, S., Mesquita, M., Villarreal, B., Fonseca, I., Francia, M. E. et al. (2018). Over-elongation of centrioles in cancer promotes centriole amplification and chromosome missegregation. *Nat. Commun.* **9**, 1258. doi:10.1038/s41467-018-03641-x
- Mennella, V., Kesztelyi, B., McDonald, K. L., Chhun, B., Kan, F., Rogers, G. C., Huang, B. and Agard, D. A. (2012). Subdiffraction-resolution fluorescence microscopy reveals a domain of the centrosome critical for pericentriolar material organization. *Nat. Cell Biol.* **14**, 1159-1168. doi:10.1038/ncb2597
- Mennella, V., Agard, D. A., Huang, B. and Pelletier, L. (2014). Amorphous no more: Subdiffraction view of the pericentriolar material architecture. *Trends Cell Biol.* **24**, 188-197. doi:10.1016/j.tcb.2013.10.001
- Moritz, M., Braunfeld, M. B., Sedat, J. W., Alberts, B. and Agard, D. A. (1995). Microtubule nucleation by γ -tubulin-containing rings in the centrosome. *Nature* **378**, 638-640. doi:10.1038/378638a0
- Nanjundappa, R., Kong, D., Shim, K., Stearns, T., Brody, S. L., Loncarek, J. and Mahjoub, M. R. (2019). Regulation of cilia abundance in multiciliated cells. *Elife* **8**, 1-23. doi:10.7554/eLife.44039
- Pazdur, R., Kudelka, A. P., Kavanagh, J. J., Cohen, P. R. and Raber, M. N. (1993). The taxoids: paclitaxel (Taxol®) and docetaxel (Taxotere®). *Cancer Treat. Rev.* **19**, 351-386. doi:10.1016/0305-7372(93)90010-0
- Petry, S., Groen, A. C., Ishihara, K., Mitchison, T. J. and Vale, R. D. (2013). Branching microtubule nucleation in *Xenopus* egg extracts mediated by augmin and TPX2. *Cell* **152**, 768-777. doi:10.1016/j.cell.2012.12.044
- Piel, M., Meyer, P., Khodjakov, A., Rieder, C. L. and Bornens, M. (2000). The respective contributions of the mother and daughter centrioles to centrosome activity and behavior in vertebrate cells. *J. Cell Biol.* **149**, 317-329. doi:10.1083/jcb.149.2.317
- Roth, D., Fitton, B. P., Chmel, N. P., Wasiluk, N. and Straube, A. (2019). Spatial positioning of EB family proteins at microtubule tips involves distinct nucleotide-dependent binding properties. *J. Cell Sci.* **132**, jcs219550. doi:10.1242/jcs.219550
- Salaycik, K. J. (2005). Quantification of microtubule nucleation, growth and dynamics in wound-edge cells. *J. Cell Sci.* **118**, 4113-4122. doi:10.1242/jcs.02531
- Salisbury, J. L., D'Assoro, A. B. and Lingle, W. L. (2004). Centrosome amplification and the origin of chromosomal instability in breast cancer. *J. Mammary Gland Biol. Neoplasia* **9**, 275-283. doi:10.1023/B:JOMG.0000048774.27697.30
- Schiff, P. B. and Horwitz, S. B. (2006). Taxol stabilizes microtubules in mouse fibroblast cells. *Proc. Natl. Acad. Sci. USA* **77**, 1561-1565. doi:10.1073/pnas.77.3.1561
- Schneeweiss, A., Sinn, H.-P., Ehemann, V., Khbeis, T., Neben, K., Krause, U., Ho, A. D., Bastert, G. and Krämer, A. (2003). Centrosomal aberrations in primary invasive breast cancer are associated with nodal status and hormone receptor expression. *Int. J. Cancer* **107**, 346-352. doi:10.1002/ijc.11408

- Semenova, I. and Rodionov, V.** (2007). Fluorescence Microscopy of Microtubules in Cultured Cells. In *Microtubule Protocols* (ed. J. Zhou), pp. 93-102. Totowa, NJ: Humana Press.
- Siegrist, S. E. and Doe, C. Q.** (2007). Microtubule-induced cortical cell polarity. *Genes Dev.* **21**, 483-496. doi:10.1101/gad.1511207
- Sonnen, K. F., Schermelleh, L., Leonhardt, H. and Nigg, E. A.** (2012). 3D-structured illumination microscopy provides novel insight into architecture of human centrosomes. *Biol. Open* **1**, 965-976. doi:10.1242/bio.20122337
- Stearns, T. and Kirschner, M.** (1994). In vitro reconstitution of centrosome assembly and function: The central role of γ -tubulin. *Cell* **76**, 623-637. doi:10.1016/0092-8674(94)90503-7
- Stepanova, T., Slemmer, J., Hoogenraad, C. C., Lansbergen, G., Dortland, B., De Zeeuw, C. I., Grosveld, F., Van Cappellen, G., Akhmanova, A. and Galjart, N.** (2003). Visualization of microtubule growth in cultured neurons via the use of EB3-GFP (end-binding protein 3-green fluorescent protein). *J. Neurosci.* **23**, 2655-2664. doi:10.1523/JNEUROSCI.23-07-02655.2003
- Straube, A. and Merdes, A.** (2007). EB3 regulates microtubule dynamics at the cell cortex and is required for myoblast elongation and fusion. *Curr. Biol.* **17**, 1318-1325. doi:10.1016/j.cub.2007.06.058
- Viola, P. and Jones, M.** (2001). Robust real-time object detection. *Int. J. Comput. Vis.* **4**, 34-47.
- Vorobjev, I. A. and Chentsov, Y. S.** (1982). Centrioles in the cell cycle. I. Epithelial cells. *J. Cell Biol.* **93**, 938-949. doi:10.1083/jcb.93.3.938
- Walker, R. A., O'Brien, E. T., Pryer, N. K., Soboeiro, M. F., Voter, W. A., Erickson, H. P. and Salmon, E. D.** (1988). Dynamic Instability of Individual Microtubules. *J. Cell Biol.* **107**, 1437-1448. doi:10.1083/jcb.107.4.1437
- Weaver, B. A.** (2014). How Taxol/paclitaxel kills cancer cells. *Mol. Biol. Cell* **25**, 2677-2681. doi:10.1091/mbc.e14-04-0916
- Woodruff, J. B., Wueseke, O. and Hyman, A.** (2014). Pericentriolar material structure and dynamics. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **5**, 369. doi:10.1098/rstb.2013.0459
- Woodruff, J. B., Wueseke, O., Viscardi, V., Mahamid, J., Zinke, A., Zanin, E., Ochoa, S. D., Decker, M., Bunkenborg, J., Baumeister, W. et al.** (2015). Regulated assembly of a supramolecular centrosome scaffold in vitro. *Science (80-.)* **348**, 808-812. doi:10.1126/science.aaa3923
- Zheng, Y., Wong, M. L., Alberts, B. and Mitchison, T.** (1995). Nucleation of microtubule assembly by a γ -tubulin-containing ring complex. *Nature* **378**, 578-583. doi:10.1038/378578a0

Figure S1

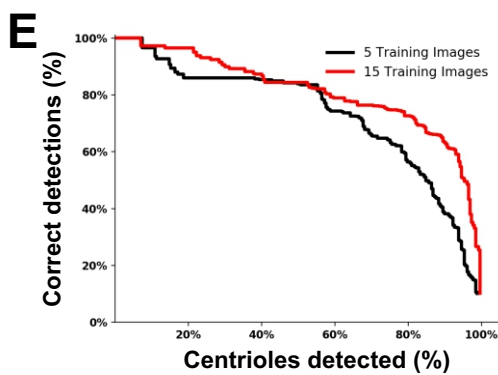
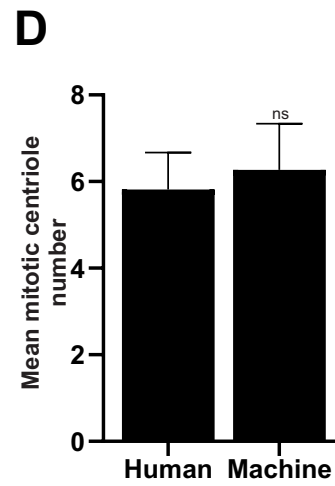
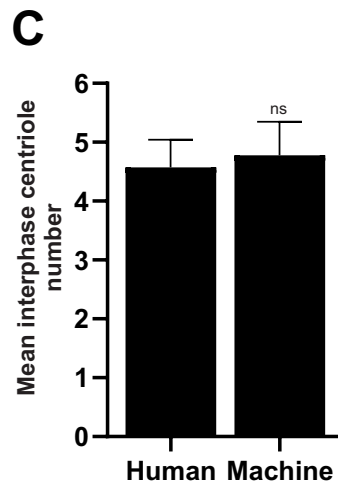
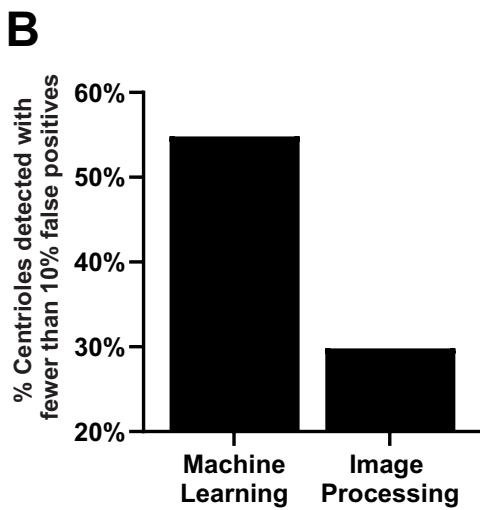
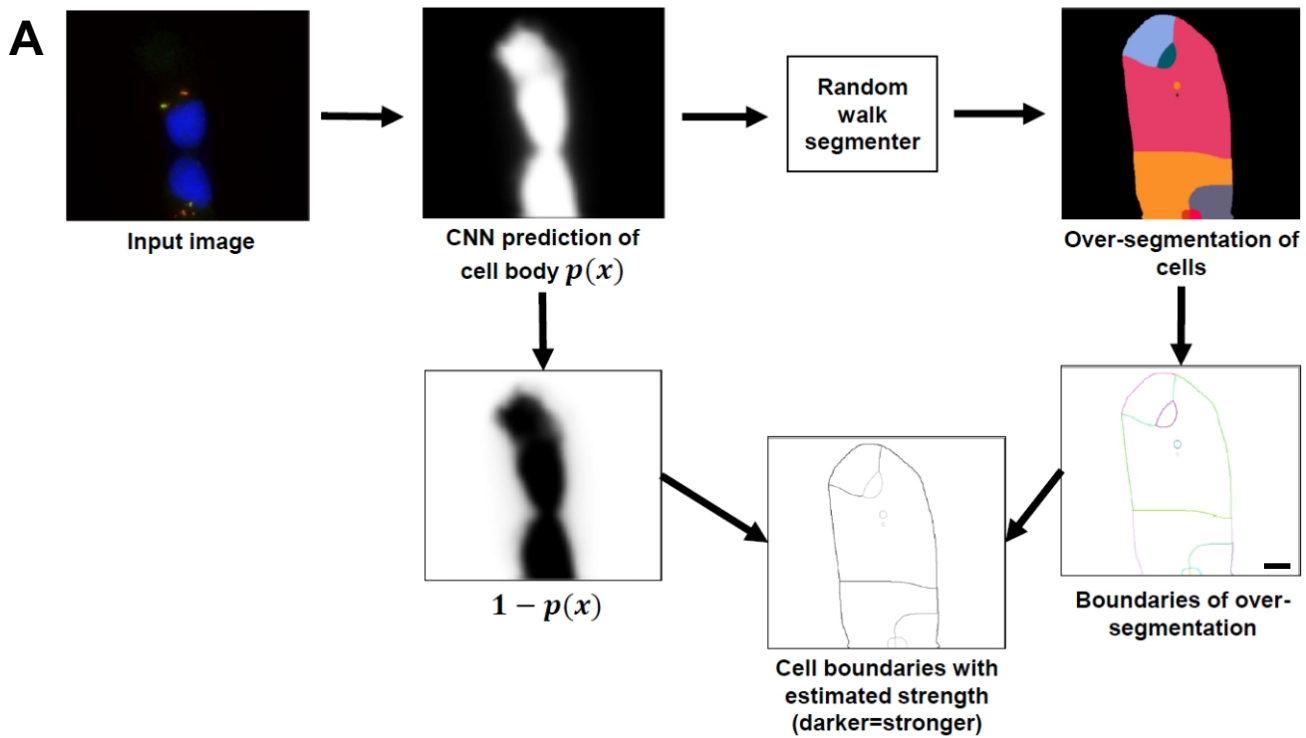


Figure S1: **(A)** Pipeline for estimating boundary strength based on the convolutional network output. The neural network estimates $p(x)$, which is the likelihood of a pixel being in the cell body (top left). The random walk segmenter uses this likelihood to over-segment cells (top right). We identify the boundary between the predicted cells (bottom left) and estimate the strength of the boundary (bottom) using $1-p(x)$ (bottom right). In the bottom panel, darker colors indicate stronger boundaries. Scale bar represent 5 μm . **(B)** Percentage of centrioles detected using machine learning versus standard image processing at score thresholds that yield fewer than 10% false positives. **(C)** Mean interphase centriole number identified by human versus machine. Mean \pm SEM. Mann-Whitney U test. ns: not significant. **(D)** Mean mitotic centriole number identified by human versus machine. Mean \pm SEM. Mann-Whitney U test. ns: not significant. **(E)** Improvement in accuracy when the centrosome detection model is trained with 15 images (red) instead of 5 images (black).

Figure S2

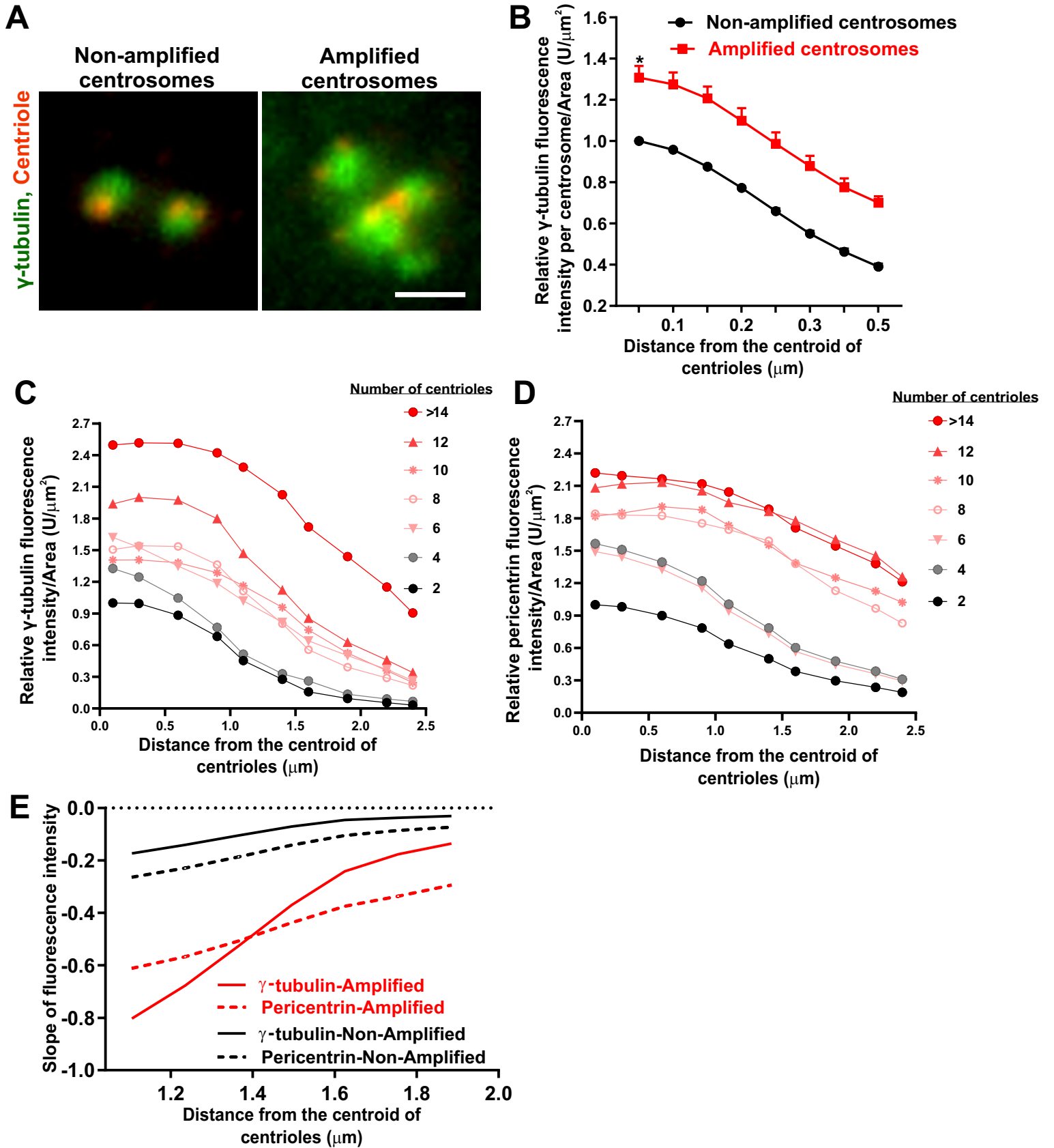


Figure S2: **(A)** Non-amplified and amplified centrosomes in S-phase arrested MDA-231 cells stained for centrioles (centrin, red) and γ -tubulin (green) Scale bars represent 1 μm . **(B)** γ -tubulin fluorescence intensity per centrosome per unit area in amplified (red) relative to non-amplified (black) MDA-231 cells obtained using human annotation (p-value: 2.7×10^{-12}). **(C)** γ -tubulin fluorescence intensity per unit area at various distances from the centroid of varying centriole numbers in non-amplified (black) and amplified (red) MDA-231 cells obtained using human annotation. **(D)** Pericentrin fluorescence intensity per unit area at various distances from the centroid of varying centriole numbers in MDA-231 cells obtained using human annotation. The slope or first derivative of fluorescence intensities of γ -tubulin and Pericentrin at non-amplified (black) and amplified centrosomes (red) outside the core pericentriolar space (between 1.0 μm to 2.0 μm). Mean \pm SEM, $\text{U}\mu\text{m}^{-2}$: Units μm^{-2} , Students t-test (paired).